



# **Universidad Nacional Mayor de San Marcos**

**Universidad del Perú. Decana de América**

**Facultad de Ingeniería de Sistemas e Informática**

**Escuela Profesional de Ingeniería de Sistemas**

## **Implementación de un modelo predictivo para incrementar la captación de seguros en una entidad financiera**

### **INFORME DE TRABAJO DE SUFICIENCIA PROFESIONAL**

**Para optar el Título Profesional de Ingeniero de Sistemas**

#### **AUTOR**

**Carolina del Carmen QUISPE CANDELA**

#### **ASESOR**

**Hugo Rafael CORDERO SÁNCHEZ**

**Lima, Perú**

**2018**



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

## Referencia bibliográfica

---

Quispe, C. (2018). *Implementación de un modelo predictivo para incrementar la captación de seguros en una entidad financiera*. Informe de Trabajo de Suficiencia Profesional para optar el título de Ingeniero de Sistemas. Escuela Profesional de Ingeniería de Sistemas, Facultad de Ingeniería de Sistemas e Informática, Universidad Nacional Mayor de San Marcos, Lima, Perú.

---

## **Hoja de metadatos complementarios**

- **Código ORCID del autor:** --
- **Código ORCID del asesor:** 0000-0003-3168-6175
- **DNI o pasaporte del autor:** 72469250
- **Grupo de investigación:** IT DATA (TI Aplicada a la Ciencia de Datos)
- **Institución que financia la investigación:** --
- **Ubicación geográfica donde se desarrolló la investigación:**  
Av. República de Panamá 3055. San Isidro (-12.093705°, -77.021076°)
- **Año o rango de años que la investigación abarcó:** 2016



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS  
FACULTAD DE INGENIERÍA DE SISTEMAS E INFORMÁTICA  
Escuela Profesional de Ingeniería de Sistemas

*Acta de Sustentación del  
Trabajo de Suficiencia Profesional*

Siendo las *17:00* horas del día *21* de noviembre del año 2018, se reunieron los docentes designados como Miembros de Jurado del Trabajo de Suficiencia Profesional, presidido por el Mg. Luza Montero Cesar (Presidente), Lic. Pantoja Collantes Jorge Santiago (Miembro) y el Ing. Cordero Sanchez Hugo Rafael (Miembro Asesor) para la sustentación del Trabajo de Suficiencia Profesional Intitulado: **"IMPLEMENTACIÓN DE UN MODELO PREDICTIVO PARA INCREMENTAR LA CAPTACIÓN DE SEGUROS EN UNA ENTIDAD FINANCIERA"**, por la Bachiller: **Quispe Candela Carolina Del Carmen**; para obtener el Título Profesional de Ingeniera de Sistemas.

Acto seguido de la exposición del Trabajo de Suficiencia Profesional, el Presidente invitó a la Bachiller a dar las respuestas a las preguntas establecida por los miembros del Jurado.

La Bachiller en el curso de sus intervenciones demostró pleno dominio del tema, al responder con acierto y fluidez a las observaciones y preguntas formuladas por los señores miembros del Jurado.

Finalmente habiéndose efectuado la calificación correspondiente por los miembros del Jurado, la Bachiller obtuvo la nota de *17*. (En letras) *Diecisiete*.

A continuación el presidente de jurados el, Mg. Luza Montero Cesar declara a la Bachiller Ingeniera de Sistemas.

Siendo las *18:00* horas, se levantó la sesión.

Presidente  
Mg. Luza Montero Cesar

Miembro  
Lic. Pantoja Collantes Jorge Santiago

Miembro Asesor  
Ing. Cordero Sanchez Hugo Rafael

## **DEDICATORIA**

Este trabajo está dedicado

A mis abuelos quienes con su amor y ejemplo

Han sido pieza fundamental en mi educación

Y han hecho la persona que soy

Desde el cielo sé que estarán contentos por este nuevo logro.

## **AGRADECIMIENTOS**

Quisiera agradecer en primer lugar a Dios y a mi familia, quienes son los que mi impulsan con su ejemplo y motivación cada día para lograr mis objetivos.

A mi asesor Hugo Cordero Sanchez, por su compromiso y acompañamiento. Quien con su guía ha hecho posible este trabajo.

A la Facultad de Ingeniería de Sistemas que a través del Programa de Titulación permite a jóvenes bachilleres que se desempeñan profesionalmente obtener el grado de ingeniero.

Gracias a todos mis compañeros que participaron en este programa, por su trabajo en equipo, colaboración y conocimientos, que nos permitieron superar satisfactoriamente los trabajos y proyectos de los cursos de los cuales hemos participado.

**UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS**  
**FACULTAD DE INGENIERÍA DE SISTEMAS E INFORMÁTICA**

**IMPLEMENTACIÓN DE UN MODELO PREDICTIVO PARA INCREMENTAR  
LA CAPTACIÓN DE SEGUROS EN UNA ENTIDAD FINANCIERA**

**Autor:** Quispe Candela, Carolina Del Carmen

**Asesor:** Cordero Sanchez, Hugo Rafael

**Título:** Informe de Trabajo de Suficiencia Profesional para optar el Título  
Profesional de Ingeniero de Sistemas

**Fecha:** Noviembre de 2018

---

**RESUMEN**

La entidad financiera tiene como meta incrementar las ventas del producto Seguro Protección Múltiple y además reducir el ratio de cancelación de 15% de este producto antes de los 3 meses, para poder tener margen de ganancias rentables. Después de haber hechos mejoras en los canales de ventas y oferta de valor, al no conseguir los objetivos propuestos. Se propuso como solución elaborar e implementar un modelo predictivo, el cual en base al análisis histórica de las altas de este producto nos permitirá identificar que clientes son más propensos a la venta de este producto como a mantener su permanencia de 3 meses como mínimo. Para su elaboración se utilizó la técnica de regresión logística debido a: su gran poder predictivo en problemas donde la variable a predecir es dicotómica (toma dos valores), es explicativa permitiéndonos interpretar el significado de las variables predictores y su relación con la variable a predecir y fácil de implementar. Utilizando la técnica se identificó 9 variables principales del modelo, logrando un rendimiento de 87.7%, lo cual hace aun modelo aceptable. Con estos resultados se desplegó el modelo en las campañas comerciales del 2016 orientando la venta a los clientes más propensos, logrando que alcanzáramos los objetivos: incremento de ventas a 710 pólizas mensuales, efectividad en campaña mayor a 2.7% y reducir el ratio de fuga de 15% a 8%.

**Palabras claves:** modelo predictivo, ventas, ratio de fuga/cancelación, regresión logística, campañas comerciales, efectividad



**MAJOR NATIONAL UNIVERSITY OF SAN MARCOS**  
**FACULTY OF SYSTEMS AND COMPUTER ENGINEERING**

**IMPLEMENTATION OF A PREDICTIVE MODEL TO INCREASE THE  
COLLECTION OF INSURANCE IN A FINANCIAL ENTITY**

**Author:** Quispe Candela, Carolina Del Carmen  
**Advisor:** Cordero Sanchez, Hugo Rafael  
**Title:** Professional Sufficiency Work Report for opt for the  
Professional Title of Systems Engineer  
**Date:** November de 2018

---

**ABSTRACT**

The financial institution aims to increase the sales of the Multiple Protection Insurance product and also reduce the 15% cancellation rate of this product before 3 months, in order to have profitable profit margins. After having made improvements in sales channels and value offer, by not achieving the proposed objectives. It was proposed as a solution to elaborate and implement a predictive model, which based on the historical analysis of the additions of this product will allow us to identify which customers are more prone to the sale of this product as well as to maintain their stay of at least 3 months. For its elaboration the logistic regression technique was used due to: its great predictive power in problems where the variable to predict is dichotomous (it takes two values), it is explanatory allowing us to interpret the meaning of the predictor variables and their relation with the variable to be predicted and easy to implement. Using the technique, we identified 9 main variables of the model, achieving a yield of 87.7%, which makes the model acceptable. With these results, the model was deployed in the commercial campaigns of 2016, orienting the sale to the most prone customers, achieving that we reached the objectives: sales increase to 710 monthly policies, campaign effectiveness greater than 2.7% and reducing the leak rate of 15% to 8%.

**Key words:** predictive model, sales, leakage / cancellation ratio, logistic regression, commercial campaigns, effectiveness

## TABLA DE CONTENIDO

PASTA O CARÁTULA EXTERNA .....	i
PÁGINA EN BLANCO .....	ii
CARÁTULA INTERNA.....	iii
FICHA CATALOGRÁFICA .....	iv
DEDICATORIA.....	v
AGRADECIMIENTOS.....	vi
RESUMEN .....	vii
ABSTRACT .....	viii
TABLA DE CONTENIDO .....	ix
ÍNDICE DE GRÁFICOS .....	xi
ÍNDICE DE TABLAS.....	xii
INTRODUCCIÓN.....	1
CAPÍTULO I. TRAYECTORIA PROFESIONAL.....	3
CAPÍTULO II. CONTEXTO EN QUE SE DESARROLLO LA EXPERIENCIA .....	8
2.1. EMPRESA - ACTIVIDAD QUE REALIZA .....	8
2.2. MISIÓN .....	9
2.3. VISIÓN .....	11
2.4. ORGANIZACIÓN DE LA EMPRESA.....	11
2.5. AREA, CARGO Y FUNCIONES DESEMPEÑADAS .....	13
2.6. EXPERIENCIA PROFESIONAL REALIZADA EN LA ORGANIZACIÓN .....	14
CAPÍTULO III. ACTIVIDADES DESARROLLADAS .....	15
3.1. SITUACIÓN PROBLEMÁTICA.....	15
3.1.1. DEFINICIÓN DEL PROBLEMA .....	15
3.2. SOLUCIÓN .....	17
3.2.1. OBJETIVOS.....	18
3.2.2. ALCANCE .....	19
3.2.3. ETAPAS Y METODOLOGÍAS .....	19
3.2.4. FUNDAMENTOS UTILIZADOS .....	44
3.2.5. IMPLEMENTACION DE LAS AREAS, PROCESOS, SISTEMAS Y BUENAS PRÁCTICAS.....	46
3.3. EVALUACIÓN .....	48
CAPÍTULO IV. REFLEXIÓN CRÍTICA DE LA EXPERIENCIA .....	53

CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES .....	56
5.1. CONCLUSIONES .....	56
5.2. RECOMENDACIONES.....	57
5.3. GLOSARIO .....	57
5.4. FUENTES DE INFORMACIÓN .....	58
ANEXOS.....	59

## ÍNDICE DE GRÁFICOS

Gráfico 1: Cobertura de la Entidad Financiera-Perú .....	9
Gráfico 2: Principios Corporativos- Entidad Financiera .....	10
Gráfico 3: Organigrama de la Organización.....	12
Gráfico 4: Organigrama de la Gerencia de Desarrollo Comercial .....	13
Gráfico 5: Funnel de Ventas del Seguro Protección Múltiple.....	16
Gráfico 6: Cancelaciones del Stock del Seguro PM.....	17
Gráfico 7: Ciclo de la Metodología Crisp-DM.....	20
Gráfico 8: Efectividad de la Campaña Seg PM-TLMK .....	21
Gráfico 9 Público para el Modelado.....	25
Gráfico 10 Diagrama Fuente Información Dataset.....	26
Gráfico 11 Diagrama de Flujo para Realizar el Modelo .....	28
Gráfico 12 Variable FLG_BUEN_COMER .....	30
Gráfico 13 Variable FLG_OPERACION_U12M .....	31
Gráfico 14: Variable T_CUENTA .....	32
Gráfico 15: Variable FLG_COLABORADOR .....	33
Gráfico 16: Curva ROC.....	34
Gráfico 17: Variables del Modelo .....	42
Gráfico 18: Comparación de la Efectividad Natural Vs el Modelo .....	42
Gráfico 19: Efectividad según Score .....	43
Gráfico 20: Algoritmo de Regresión Logística .....	45
Gráfico 21: Flujo de Utilización del Modelo en las Campañas.....	46
Gráfico 22: Indicadores de los Pilotos Comerciales Seg PM.....	50
Gráfico 23: Ratio de Cancelaciones de los Pilotos.....	51
Gráfico 24: Ventas Seg PM 2015 vs 2016 .....	52

## ÍNDICE DE TABLAS

Tabla 1: Variables internas de entidad Financiera.....	23
Tabla 2: Variables del SSFF.....	23
Tabla 3: Variables socio demográficas.....	24
Tabla 4: Base de Datos- Fuente de Información .....	27
Tabla 5: Comparación de Rendimiento de Algoritmos .....	35
Tabla 6: Resultados de Iteraciones del Modelado .....	38
Tabla 7: Coeficientes del Modelo de Regresión Logística.....	41

## INTRODUCCIÓN

El presente Informe Profesional desarrolla la implementación de un modelo predictivo para la captación del Seguro Protección Múltiple, la cual está enfocado en incrementar las ventas y reducir el ratio de 15% de la fuga de este producto a los 3 meses post venta.

El modelo predictivo se desarrolló con la técnica de regresión logística debido a: su gran poder predictivo en problemas donde la variable a predecir es dicotómica (toma dos valores), es explicativa permitiéndonos interpretar el significado de las variables predictores y su relación con la variable a predecir y fácil de implementar. Logrando un rendimiento del modelo de 87.7%, lo cual indica que es un modelo muy bueno.

El uso del modelo predictivo en la organización en el año 2016 nos permitió resolver la problemática presentada en la organización y llegar a los indicadores establecidos en la organización, obteniendo más de 710 ventas en promedio con una efectividad en campaña mayor a 2.7% y reduciendo el ratio de fuga de 15% a 8%.

Con los buenos resultados de esta solución el área de Inteligencia Comercial en coordinación con el Área de Seguros, impulsaron el desarrollo de nuevos modelos para otros productos de seguros para mejorar las campañas comerciales.

El presente informe se desarrolla en base a los siguientes capítulos:

En el Capítulo I, desarrollaré mi trayectoria profesional, la cual refleja mi experiencia adquirida que me permitió participar en todas las etapas de la elaboración del modelo predictivo.

En el Capítulo II, desarrollaré el contexto en el que se desarrolló la experiencia, describiendo a la Entidad Financiera, su visión y misión, organización, área cargo y funciones desempeñadas, así como la experiencia profesional realizada en la organización.

En el Capítulo III, explicaré el problema presentado, los objetivos y alcance, las etapas y metodología empleadas, los fundamentos utilizados e implementación, así como la evaluación de la solución.

En el Capítulo IV, expondré la reflexión crítica de la experiencia.

En el Capítulo V, desarrollaré las conclusiones y recomendaciones del presente Informe Profesional.

## CAPÍTULO I. TRAYECTORIA PROFESIONAL

La autora es Bachiller en Ingeniería de Sistemas e Informática quien posee una especialización en Big Data y Analytics. Presenta experiencia en diferentes puestos del área de inteligencia de negocios o comercial en el sector bancario, donde a través de la tecnología y la analítica aplicadas al marketing ayuda al negocio en sus estrategias comerciales.

A continuación, se detalla la trayectoria profesional:

Experiencia Profesional	
Mayo 2018 - Actualidad	<p><b>Cargo: Especialista 2 en Campañas Dirigidas</b></p> <p><b>Área:</b> Inteligencia Comercial - BBVA Continental</p> <p><b>Actividades y Logros:</b></p> <ul style="list-style-type: none"><li>▪ Optimización del proceso de campañas del segmento Empresas y Negocios.</li><li>▪ Implementación y despliegue de las campañas del segmento Negocios a los canales de venta de los siguientes productos (TKT, Pcom, Línea Pcom, Comex, Pasivos y Servicios).</li><li>▪ Automatización de proceso de despliegue de email por salesforce de las campañas de Personas Jurídicas.</li></ul>



	<ul style="list-style-type: none"> <li>▪ Análisis de vinculación de la cartera de clientes del segmento negocio para determinar las adecuadas acciones comerciales de vinculación, fidelización y retención de clientes.</li> <li>▪ Análisis de los clientes Attrition donde se determina las principales acciones comerciales para evitar la fuga de clientes.</li> </ul>
Agosto 2015- Abril 2018	<p><b>Cargo: Especialista 3 en Campañas Dirigidas</b></p> <p><b>Área:</b> Inteligencia Comercial - BBVA Continental</p> <p><b>Actividades y Logros:</b></p> <ul style="list-style-type: none"> <li>▪ Generación y distribución de campañas con visión cliente.</li> <li>▪ Análisis de las campañas comerciales del segmento personas naturales enfocados de los siguientes productos: Vehicular, Préstamo de Libre disponibilidad, Hipotecario, Seguros y FFMM, con el objetivo de determinar los públicos objetivos que maximicen la efectividad.</li> <li>▪ Mejora del proceso de distribución a los canales Digitales (Bnet, Agentes, Atm, Segmentador de Colas, SMS) con la participación del  Equipo de ingeniería logrando tener el primer día útil las campañas desplegadas.</li> <li>▪ Creación de Reporterías y definición de Kpis de las campañas Digitales (Adelanto de Sueldo, Pat, Seguro PT, Fondos Mutuos, etc.)</li> <li>▪ Participación en los proyectos para el desarrollo de la herramienta CRM del Banco (PIC/Datazo)</li> </ul>
Febrero 2015- Agosto 2015	<p><b>Cargo: Especialista en Conocimiento de Cliente</b></p>

	<p><b>Área:</b> Inteligencia Comercial - BBVA Continental</p> <ul style="list-style-type: none"> <li>▪ Análisis de los clientes y no clientes Bancarizados para buscar oportunidades de penetración de Mercado.</li> <li>▪ Actualización de los cubos de Información del Banco del segmento de Personas Naturales</li> <li>▪ Elaboración de un sistema local para integrar las solicitudes de información de las unidades del Banco al área de Inteligencia comercial mediante el uso de google forms y java scripts</li> <li>▪ Generación de reportes para las áreas del Banco como: Segmento, productos y canales</li> <li>▪ Elaboración de cuadros de mando para el seguimiento de la cartera y vinculación de los clientes del segmento negocio.</li> </ul>
--	---

Formación Académica	
Marzo 2010- Febrero 2015	<p><b>Universidad Nacional Mayor de San Marcos – Facultad de Ingeniería de Sistemas e Informática</b></p> <p>Bachiller en Ingeniería de Sistemas</p> <p>Obtención de Grado:</p> <p>Logro: Reconocimiento Académico 2014 por el primer lugar en Rendimiento académico en la EAP de Ingeniería de Sistemas.</p> <p>Promoción 2010 – Otorgado por UNMSM</p>

Marzo Diciembre 2009	2005- 2009	<b>I.E.P Mixto San José</b>  Educación Secundaria  Logro: Primer Puesto Promoción 2009
-------------------------	---------------	--

Curso y Capacitaciones		
2018-2018		<b>Conferencia Big Data Summit 2018</b>
2017 -2018		<b>Uni – Centro de Tecnologías de Información y Comunicaciones</b>  Programa de Especialización en Big Data y Analytics (120 horas)
2017 -2017		<b>Conferencia de Big Data Summit 2017</b>
2015 -2016		<b>Cibertec</b>  Business Intelligence Application Developer SQL Server 2012 (80 horas)
2014-2014		<b>IPAE</b>  Cade Universitario 2014
2014 – 2014		<b>Universidad Nacional de Ingeniería</b>  Curso Java Aplicaciones Web
2011-2011		<b>Universidad Nacional Mayor de San Marcos</b>  Curso de Herramientas Informáticas
2007-2007		<b>Universidad Nacional de Ingeniería</b>  Corel Draw 12.0 – Designer
Otros Conocimientos		

Base de Datos:	Oracle 11 g, Mysql, Sql Server 2012
Lenguaje de Programación:	Java, C++, R, Scala, Spark SQL
Herramienta Modelado	IBM Rational
Otros:	Manejo de Ireport 5.1.0, Office, Visual Studio

## **CAPÍTULO II. CONTEXTO EN QUE SE DESARROLLO LA EXPERIENCIA**

### **2.1. EMPRESA - ACTIVIDAD QUE REALIZA**

La Entidad Financiera es una sólida empresa de gran prestigio en el ámbito nacional e internacional. La cual desarrolla sus actividades a través de una red nacional conformada por 370 oficinas distribuidas en todo el país (Ver Gráfico 1), ubicada en el distrito de San Isidro, ciudad de Lima, capital del Perú. La Entidad Financiera es una sociedad anónima constituida en 1951, autorizada a operar por la Superintendencia de Banca, Seguros y AFP (SBS).

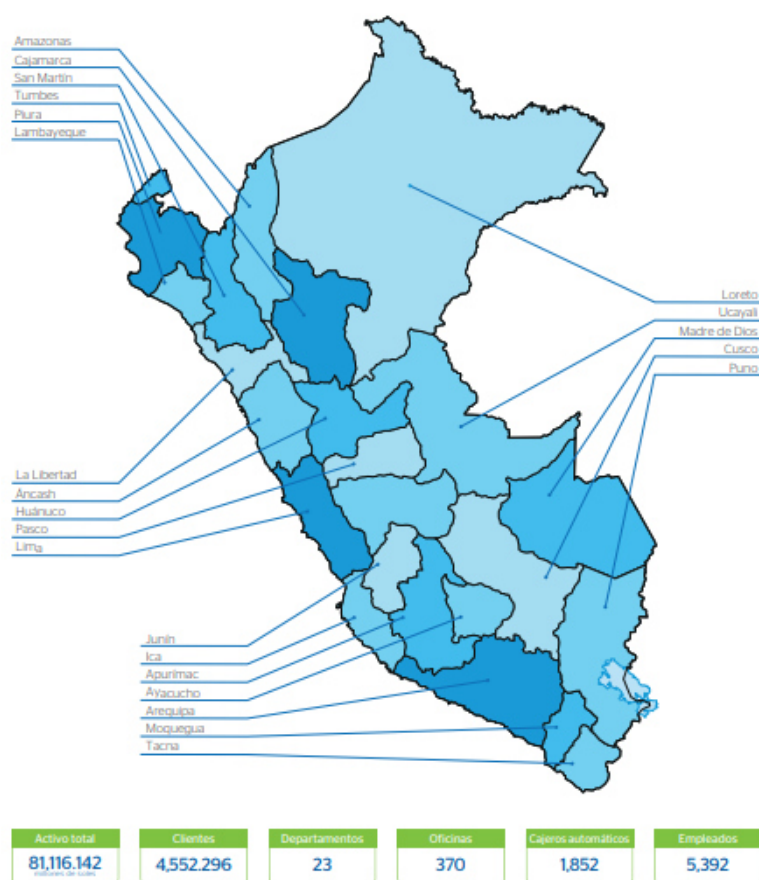


Gráfico 1: Cobertura de la Entidad Financiera-Perú (Fuente: Entidad Bancaria)

## 2.2. MISIÓN

La Entidad Financiera tiene como misión “Ofrecer las mejores soluciones, productos y servicios financieros, con sencillez y responsabilidad, a sus diferentes grupos de interés: clientes, empleados, accionistas, proveedores y sociedad”.

Para esta entidad sencillez significa agilidad, cercanía, accesibilidad y claridad en el lenguaje. Y cuando se habla de responsabilidad se expresa ese enfoque de largo plazo que caracteriza a la organización, a través de una relación equilibrada con los clientes y del apoyo al desarrollo de la sociedad.

Con el propósito de alcanzar estos objetivos, se ha definido 7 principios corporativos que tienen que ver con las personas, la ética y la innovación (Ver Gráfico 2).



Gráfico 2: Principios Corporativos- Entidad Financiera (Fuente: Entidad Bancaria)

## 2.3. VISIÓN

“La búsqueda permanente de un mejor futuro para las personas”, la cual desarrolla bajo 3 pilares:

- **Integridad**, como manifestación de la ética en sus actuaciones y en todas las relaciones con diferentes grupos de interés.
- **Prudencia**, como principio de cautela en la toma de riesgos.
- **Transparencia**, como máxima para ofrecer un acceso a información clara y veraz dentro de los límites de la legalidad.

## 2.4. ORGANIZACIÓN DE LA EMPRESA

Organigrama de la Empresa en el periodo 2016 donde se llevó a cabo el desarrollo del modelo predictivo para la captación de seguros (Ver Gráfico 3).



## Organigrama y estructura de negocio

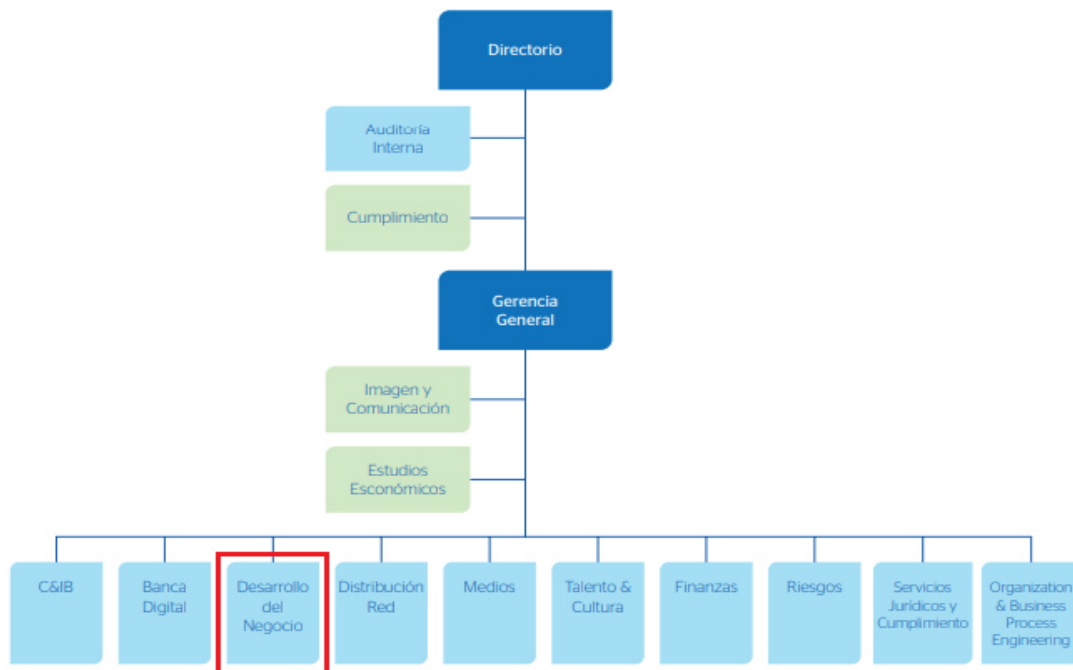


Gráfico 3: Organigrama de la Organización (Fuente: Elaboración Propia)

Dentro de la unidad de Desarrollo de Negocio, se encuentra la Gerencia de Desarrollo Comercial, la cual está constituida de la siguiente manera: (Ver Gráfico 4)

## Organigrama de Desarrollo Comercial

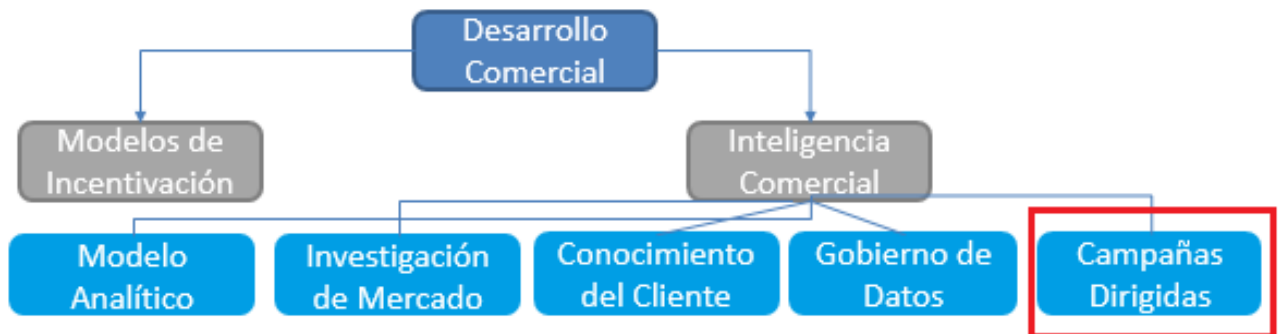


Gráfico 4: Organigrama de la Gerencia de Desarrollo Comercial (Fuente: Elaboración Propia)

Dentro de esta Gerencia se encuentra el equipo de Campañas Dirigidas, el cual tiene como función la definición, generación y delivery de las campañas comerciales de la entidad Financiera.

### 2.5. AREA, CARGO Y FUNCIONES DESEMPEÑADAS

El autor de este Informe de Experiencia Profesional se desempeñó como Especialista de Conocimiento del Cliente y Campañas Dirigidas dentro la entidad Financiera. En el periodo 2016 durante el desarrollo del proyecto descrito se desempeñó como especialista de campañas dirigidas, donde las principales funciones es realizar análisis, el targeting en base a reglas comerciales y el delivery de las campañas comerciales del segmento personas naturales en todos los canales (Fuerza de Venta- Telemarketing, Red de oficinas, canales digitales).

## 2.6. EXPERIENCIA PROFESIONAL REALIZADA EN LA ORGANIZACIÓN

### Principales Funciones:

- Generación y distribución de campañas con visión cliente.
- Análisis de las campañas comerciales del segmento personas naturales enfocados de los siguientes productos: Vehicular, Préstamo de Libre disponibilidad, Hipotecario, **Seguros** y Fondos Mutuos, con el objetivo de determinar los públicos objetivos que maximicen la efectividad.
- Mejora del proceso de distribución a los canales Digitales (Bnet, Agentes, ATM Segmentador de Colas, SMS) con la participación del Equipo de ingeniería logrando tener el primer día útil las campañas desplegadas.
- Creación de Reporterías y definición de KPI's de las campañas Digitales (Adelanto de Sueldo, Pat, Seguro Protección de Tarjeta, Fondos Mutuos, etc.)
- Participación en los proyectos para el desarrollo de la herramienta CRM del Banco (PIC/Datazo)

## CAPÍTULO III. ACTIVIDADES DESARROLLADAS

### 3.1. SITUACIÓN PROBLEMÁTICA

#### 3.1.1. DEFINICIÓN DEL PROBLEMA

En el año 2016 en la Entidad Financiera se realizó un análisis de las ventas del producto “seguro protección múltiple”, donde se analizó 2 factores: Ventas en campañas y el comportamiento del stock del producto.

De las **Ventas en Campaña**, se detectó el problema del *decrecimiento en ventas del producto*, como se puede observar en el Gráfico 5 en los últimos 4 meses de ventas (Periodo Sep –Dic 15) la efectividad había decrecido de 2.7% a 2.3%, teniendo ventas menores al promedio anual de 660 altas mensuales, lo cual impactaría directamente en el indicador anual del 2016, el cual establece tener ventas de 700 pólizas mensuales y una efectividad en campaña mayor igual a 2.8%.

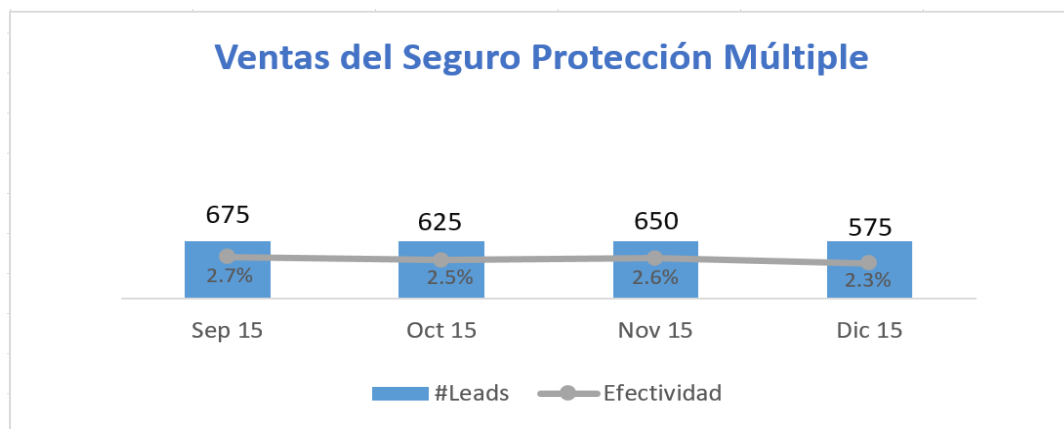


Gráfico 5: Funnel de Ventas del Seguro Protección Múltiple (Fuente: Elaboración Propia)

A pesar de que en estos 4 meses el equipo de Seguros incorporó más personal en el call center dedicado a la venta de este producto e implantó 2 acciones como política de gestión: recorrer la base 3 veces para incrementar la contactabilidad y un nuevo speech de gestión no logró incrementar las ventas ni la efectividad que son métricas exigidas por la entidad.

Del análisis del **comportamiento del Stock**, se detectó un segundo problema:

Existe un ratio de *fuga de 15% de clientes*, lo que representa a 126 clientes en promedio (Ver Gráfico 6), a los 3 meses después de haber realizado la venta originando pérdidas financieras ya que el negocio necesita como mínimo 3 meses para recuperar la inversión realizada en la gestión de la venta del producto como: pago de personal, pago de comisiones y gastos en refuerzo con el uso de email para reforzar la venta.

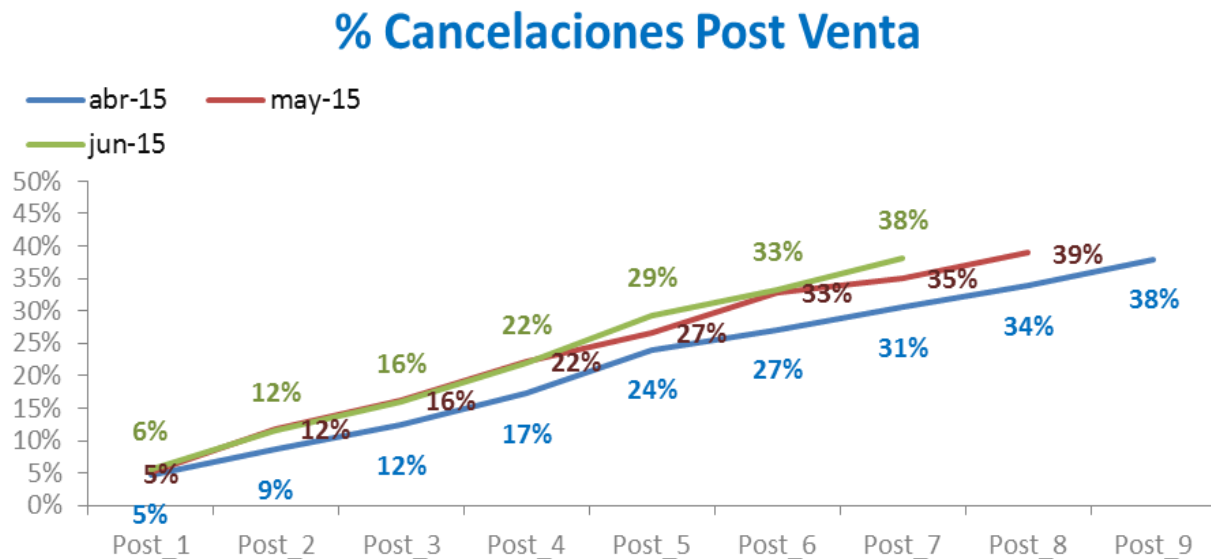


Gráfico 6: Cancelaciones del Stock del Seguro PM (Fuente: Elaboración Propia)

Del análisis respectivo se concluyó que existe 2 principales problemas que se deben abordar: decrecimiento de ventas y el ratio de fuga de clientes a los 3 meses post venta.

### 3.2. SOLUCIÓN

Como solución el equipo de Inteligencia Comercial propuso implementar un modelo predictivo para la compra del seguro protección múltiple (con permanencia al menos de 3 meses) usando técnicas de aprendizaje automático.

Un modelo predictivo se construye en base al análisis histórico de los datos que describen el problema y con el empleo del algoritmo adecuado nos permite conocer las relaciones entre las variables de entrada y salida, obteniendo un mejor conocimiento del cliente y obtener así la capacidad de poder predecir sus acciones futuras.

Con la aplicación del modelo entrenado sobre nuevos datos nos podrá identificar los clientes con mayor probabilidad a aceptar el producto de Seguro de Protección Múltiple y además que permanezcan con nuestra entidad un plazo mínimo de 3 meses. Con el empleo de esta solución podemos direccionar nuestras estrategias comerciales a estos segmentos de clientes potenciales, orientando a nuestros canales de ventas a la priorización de gestión, logrando incrementar las ventas y reducir el ratio de fuga del 15%.

### **3.2.1. OBJETIVOS**

#### **OBJETIVO GENERAL**

Implementar el modelo predictivo de captación de seguro de Protección Múltiple para incrementar las ventas y conseguir la permanencia de las altas seguro durante un periodo mínimo de 3 meses

#### **OBJETIVOS ESPECIFICOS**

- Identificar las principales variables (al menos 5) que determinen el perfil de un cliente de la entidad financiera con una alta probabilidad a la compra de Seguro Protección Múltiple

- Implementar el modelo predictivo para la compra del seguro protección múltiple.
- Probar la solución en una campaña comercial durante 2 meses para validar los resultados esperados.
- Incrementar las ventas con la utilización del modelo predictivo
- Reducir el ratio de fuga de 15% a los 3 meses post venta

### **3.2.2. ALCANCE**

El presente informe brinda información de los pasos realizados para el desarrollo e implementación del modelo predictivo para la compra del seguro protección múltiple, así como las mejoras que originó su utilización en su despliegue en las campañas comerciales.

### **3.2.3. ETAPAS Y METODOLOGÍAS**

Para el desarrollo del modelo predictivo se utilizó como guía la metodología CRISP-DM, la cual está compuesta de 6 fases. (Ver gráfico7)



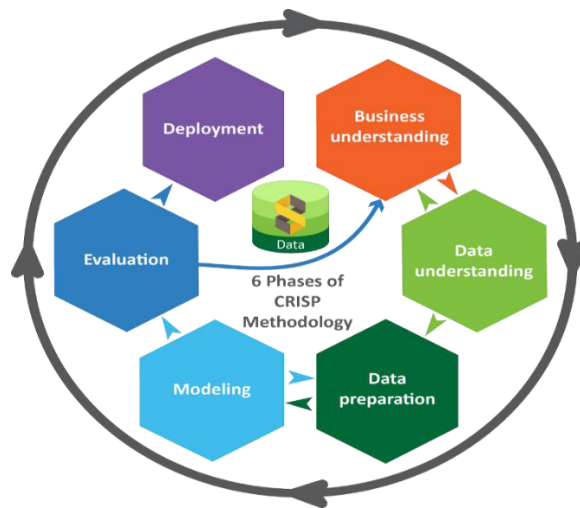


Gráfico 7: Ciclo de la Metodología Crisp-DM (Fuente: Crisp-DM)

A continuación, se describirá las actividades realizadas en las etapas aplicadas para el proyecto de la metodología empleada:

## 1. Entendimiento del Negocio

### Objetivos del Negocio:

El área de producto de Seguros de la entidad Financiera tiene como KPI incrementar las ventas del producto Seguro PM a más de 700 altas mensuales realizadas en el canal de Telemarketing.

### Evaluación de la situación:

En el análisis realizado sobre las campañas de TLMK del seguro de PM en el periodo Septiembre – diciembre 2015, se observa que la efectividad había decrecido de 2.7% a 2.3%. Se entiende por efectividad al número de altas entre los prospectos enviados al canal para la venta.

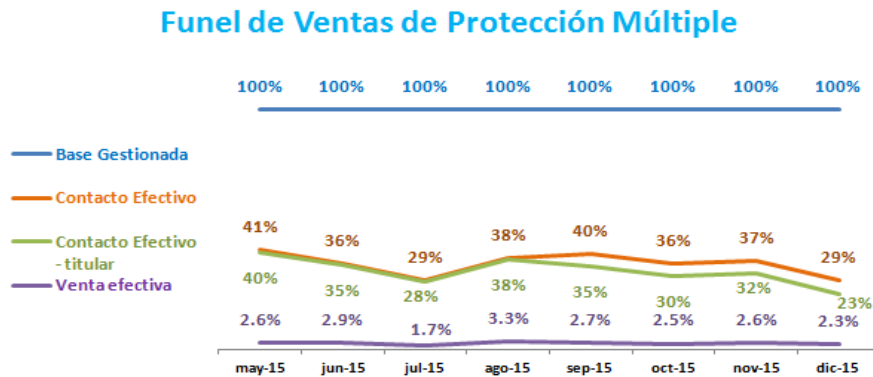


Gráfico 8: Efectividad de la Campaña Seg PM-TLMK (Fuente: Elaboración Propia)

Adicional se detectó que en el stock del producto tiene una tasa de fuga de 15% se realiza a los 3 meses post venta.

### Objetivos de la Minería de Datos:

Realizar un modelo predictivo que permita identificar las principales variables que determinen el perfil de un cliente de la entidad Financiera con una alta probabilidad a la compra de Protección Múltiple (permanencia al menos 3 meses)

### Criterios de Éxito:

El criterio de éxito es que el modelo implementado en la campaña de seguro comercial tenga mayor efectividad (supere el 2.7%) que una campaña solo basada en las reglas comerciales definidas por el negocio en el año 2015. Para ello se ejecutará un piloto por 2 meses donde el indicador que prevalecerá es la efectividad.

Como segundo indicador se revisará que la cancelación de las altas realizadas con la implementación del modelo sea menor al 15%(3 meses post venta).

### **Generación del Plan del Proyecto:**

Se estableció un tiempo de 2 meses para la elaboración del modelo, donde el primer mes se enfocó en reuniones semanales con el área de negocio para poder comprender las variables identificadas en el proceso de modelado y tener una retroalimentación del perfil del cliente que compra el seguro PM y conocer el producto.

A partir del 3 mes se planificó la generación de pilotos comerciales para poder corroborar los criterios de éxito.

## **2. Comprensión de los Datos**

Se recolectó la información disponible de los clientes de la entidad Financiera, donde se clasificó a 225 variables en 3 tipos:

- **Variables Internas de la Entidad Financiera:** Información de las actividades vinculadas al cliente con la entidad financiera.

Variable
Tenencias de Productos
Segmento del cliente
Vínculo con la entidad Financiera
Transacciones en los canales

Tabla 1: Variables internas de entidad Financiera (Fuente: Elaboración Propia)

- **Variables de endeudamiento del SSFF:** Información crediticia de saldos distribuida por la SBS a las entidades financieras.

Variable
Saldos como PN en Banco1
Saldos como PN en Banco2
Saldos por producto por entidad
Nivel de Vinculación con otros bancos

Tabla 2: Variables del SSFF (Fuente: Elaboración Propia)

- **Variable Socio demográficas:** Indicadores sociales, económicos y demográficos que permiten segmentar la población en grupos homogéneos

Variable
Sexo
Edad
Ubigeo
Nivel de Ingreso

Tabla 3: Variables socio demográficas (Fuente: Elaboración Propia)

### 3. Preparación de los Datos

En esta fase construimos el dataset que utilizaremos para el modelado, definiendo las fuentes de información de las variables y consolidándolo en una sola matriz de información.

El dataset que contiene las variables para el aprendizaje del modelo se basa en información de los clientes que han sido enviados a gestionar para la venta de este producto (historia de 6 meses de campañas) y se miden los resultados (variable target) con las altas de clientes con al menos permanencia de 3 meses post venta. (Ver Gráfico 9)

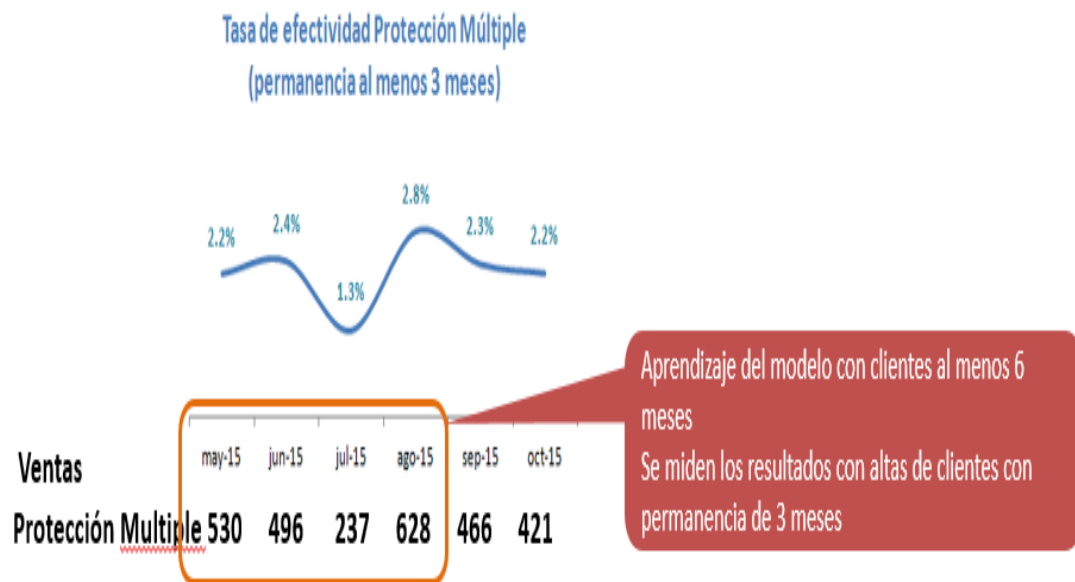


Gráfico 9 Público para el Modelado (Fuente: Elaboración Propia)

En el Gráfico 10, se esquematiza las fuentes de información para el dataset usado en el modelado:

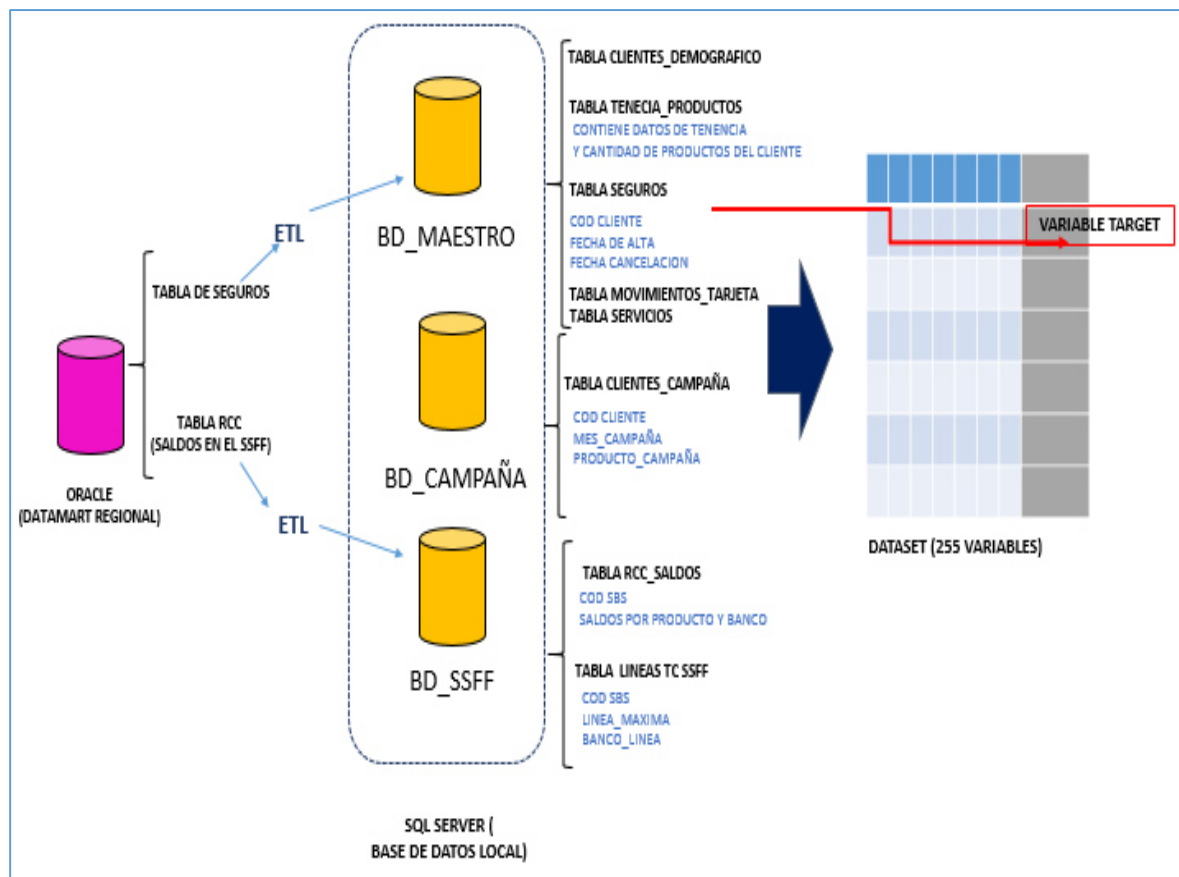


Gráfico 10 Diagrama Fuente Información Dataset (Fuente: Elaboración Propia)

Donde tenemos como:

- **Repositorio de Fuentes de Información:**



	<p><b>Base de Datos Local:</b> Contiene la información para la analítica del área de Inteligencia Comercial de Perú. La base de Datos es SQL Server 2014 y se encuentra físicamente dentro de la entidad Financiera.</p>
	<p><b>Data Warehouse:</b> Conocido como LAR (Laboratorio Regional Analítico), contiene la información de los clientes para el desarrollo analítico de la Región de las áreas de Inteligencia Comercial, teniendo como ubicación física a México. La Base de Datos es Oracle.</p> <p>Desde Perú accedemos con una conexión cliente con el PL/SQL Developer.</p>

Tabla 4: Base de Datos- Fuente de Información (Fuente: Elaboración Propia)

Dentro de la Base de Datos Local, existen 3 Bases de Datos:

- ✓ BD\_MAESTRO: Contiene tablas relacionadas con la información demográfica de los clientes, productos y servicios adquiridos por el cliente con la entidad financiera.
- ✓ BD\_CAMPAÑA: Información de los clientes y atributos de oferta comercial enviados por cada campaña.
- ✓ BD\_SSFF: Información de los saldos y productos otorgado por la SBS a todas las entidades financieras.



- **Los procesos de ETL:** Se realizan desde el Data Warehouse hacia la Base de Datos Local, utilizando la herramienta Microsoft SSIS.

#### 4. Modelado

En esta etapa para el modelado se usó la herramienta **TIMI Modeler** en la cual trabajamos las variables construidas en el dataset en la etapa anterior e identificamos las principales variables que permitan determinar el perfil de compra y permanencia de un periodo mínimo de 3 meses.

Para el modelado se siguió los siguientes pasos descritos en la Gráfica 11.

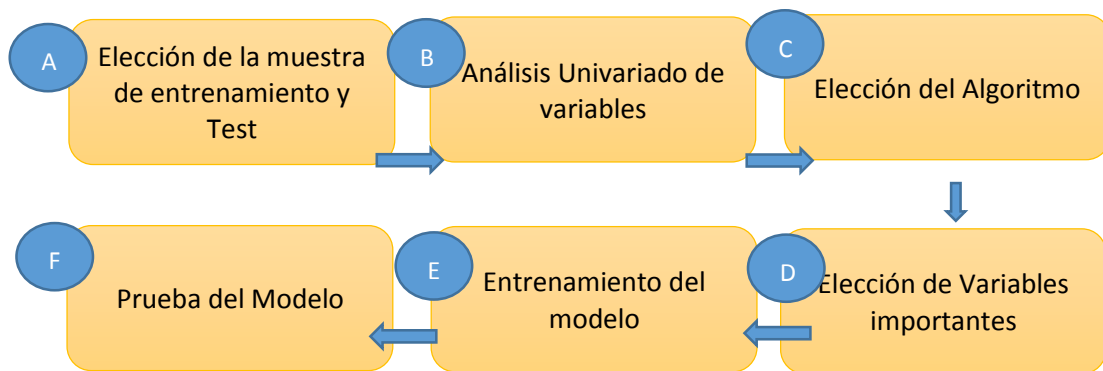


Gráfico 11 Diagrama de Flujo para Realizar el Modelo (Fuente: Elaboración Propia)

Para la realización del modelo se trabajó de la mano con un especialista en estadística.

A continuación, detallaremos lo realizado en cada paso:

### ***A. Elección de la muestra de Entrenamiento y Test***

Para la **muestra de entrenamiento** se consideró la información de los clientes de 6 meses que fueron gestionados en la campaña de Seg. PM (mayo 15 – octubre 15), lo cual equivale a 99 K registros de análisis.

Para la **muestra de test** se consideró la información de los clientes gestionados en los meses de Oct 15-Nov 15.

Como **variable Target**, clientes que han aceptado el producto y han permanecido con el producto a menos 3 meses se identificó 3,3 K registros, lo cual representa el 2% de la base de entrenamiento. El tamaño de muestra de variable target es muy inferior al 15% requerido para este tipo de problemas.

### ***B. Análisis Univariado***

Se realizó el análisis Univariado, lo cual nos permite conocer el comportamiento de cada variable independiente. Nos interesa conocer el comportamiento de cada variable (Frecuencia) y detectar los missing (datos vacíos o nulos) para poder depurarlos.

Ejemplos de Variables con alta concentración de Missing

**Variable FLG\_BUEN\_COMER:** Nos identifica el nivel de consumo en el rubro de restaurantes realizadas por tarjetas de crédito o débito.

Se puede observar que el 55% de la base para esta variable está compuesta por datos missing.

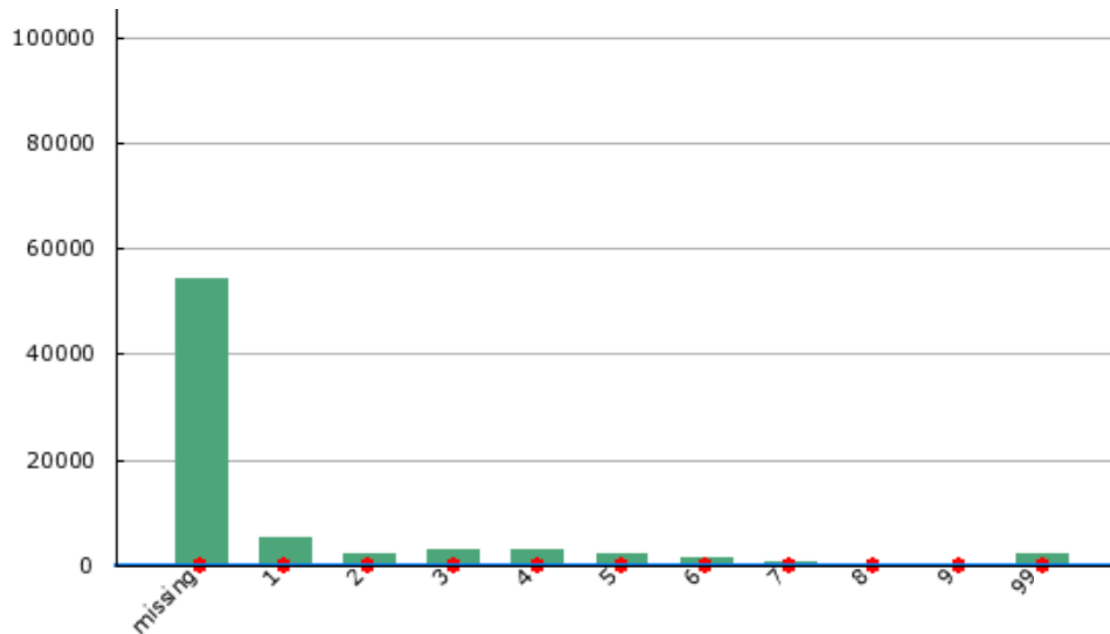


Gráfico 12 Variable FLG\_BUEN\_COMER (Fuente: Elaboración Propia)

**Variable FLG\_OPERACION\_U12M:** Si el cliente ha usado sus tarjetas en el periodo de 12 Meses.

Se puede observar que el 66% de la base para esta variable está compuesta por datos missing. (Ver Gráfico 13)

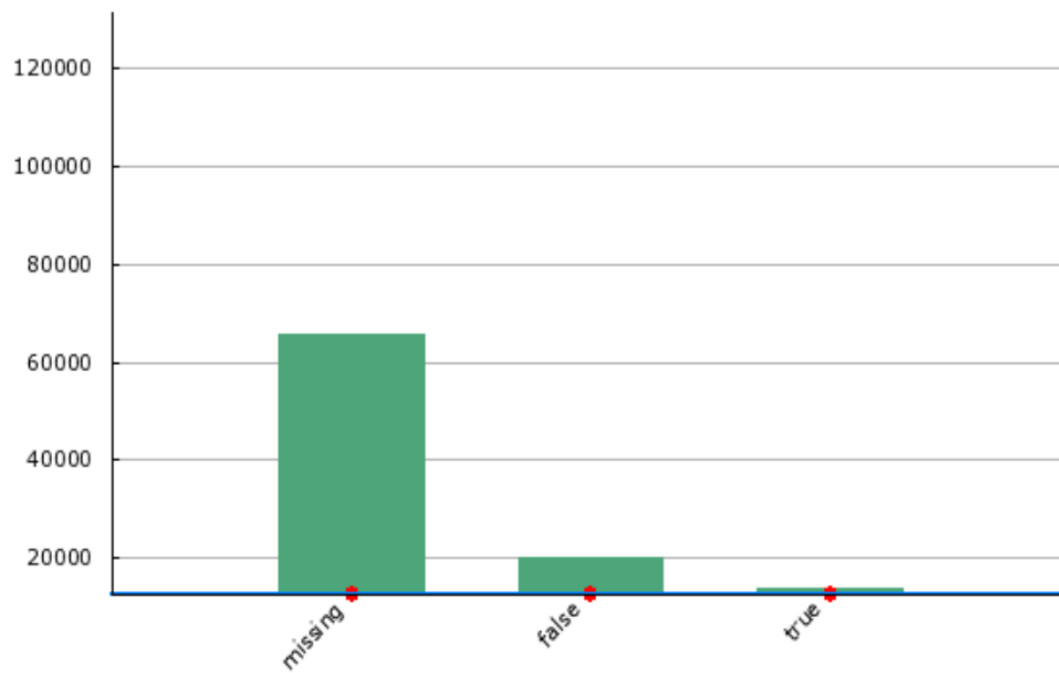


Gráfico 13 Variable FLG\_OPERACION\_U12M (Fuente: Elaboración Propia)

**Variable T\_CUENTA:** Tiene un pasivo con la entidad financiera.

El 100% tiene un pasivo, al no haber variedad en la variable se tiene que excluir del análisis para no distorsionar el modelado. (Ver Gráfico 14)

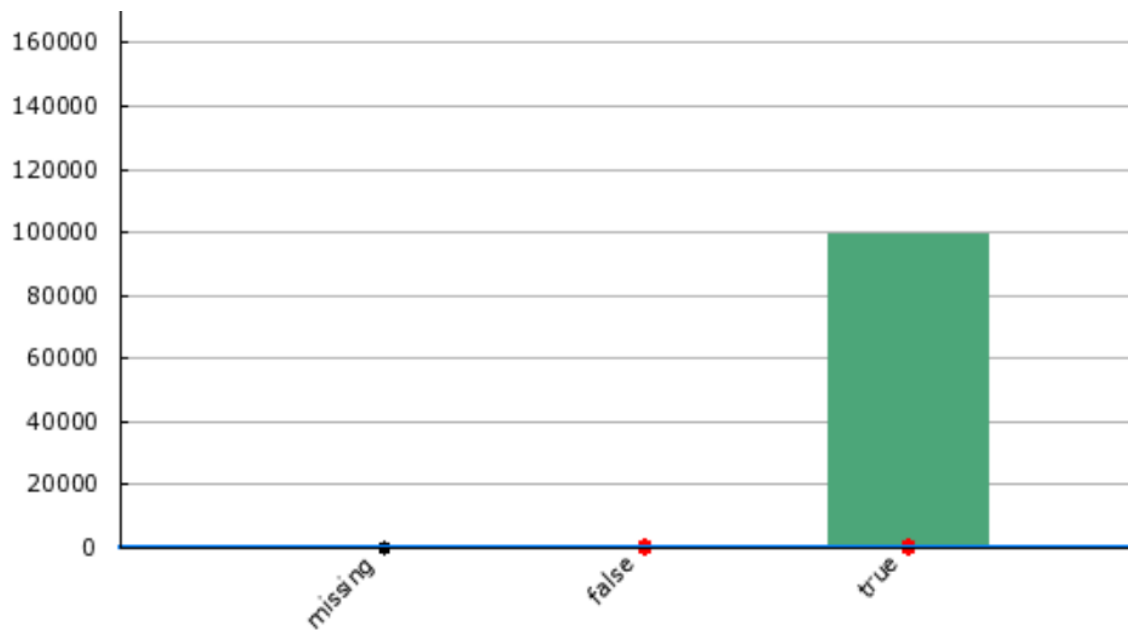


Gráfico 14: Variable T\_CUENTA (Fuente: Elaboración Propia)

**Variable FLG\_COLABORADOR:** Es un empleado de la entidad Financiera.

El 100% no es un colaborador de la entidad Financiera. Al no ver variedad en la variable debemos descartarla del entrenamiento del modelo (Ver Gráfico 15)

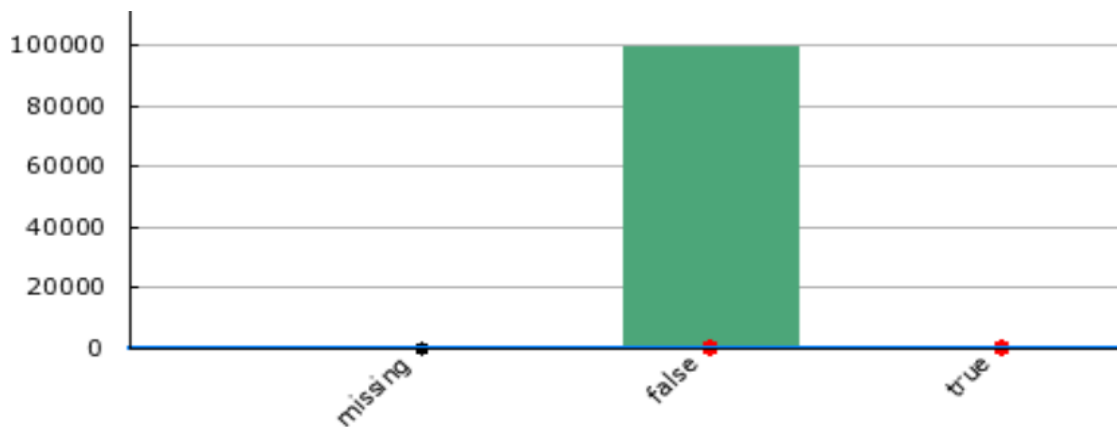


Gráfico 15: Variable FLG\_COLABORADOR (Fuente: Elaboración Propia)

### *C. Elección del Algoritmo*

Como siguiente paso escogemos el algoritmo que mejor se ajuste al problema. Ya que estamos en un problema de clasificación, debido a que la variable objetivo es **dicotómica** (solo acepta dos valores): “1” si aceptó el producto/se mantuvo 3 meses con la entidad y 0 si no tomo el producto.

Por la experiencia con otros trabajos de modelado para captación de productos en la entidad financiera, se recomienda elegir entre un árbol jerárquico o una regresión logística ya que se puede tener un mayor entendimiento del modelo y la explicación de los resultados.

Para ello se aplicó ambos algoritmos sobre la base de entrenamiento y se eligió al que tenía mayor **AUC**.

**El AUC (Área bajo la curva ROC):** Es un indicador que nos permite medir el rendimiento de un modelo en base a que tan bien puede discriminar los valores positivos de los negativos. (Hanley J.A. & McNeil B.J., 1982)

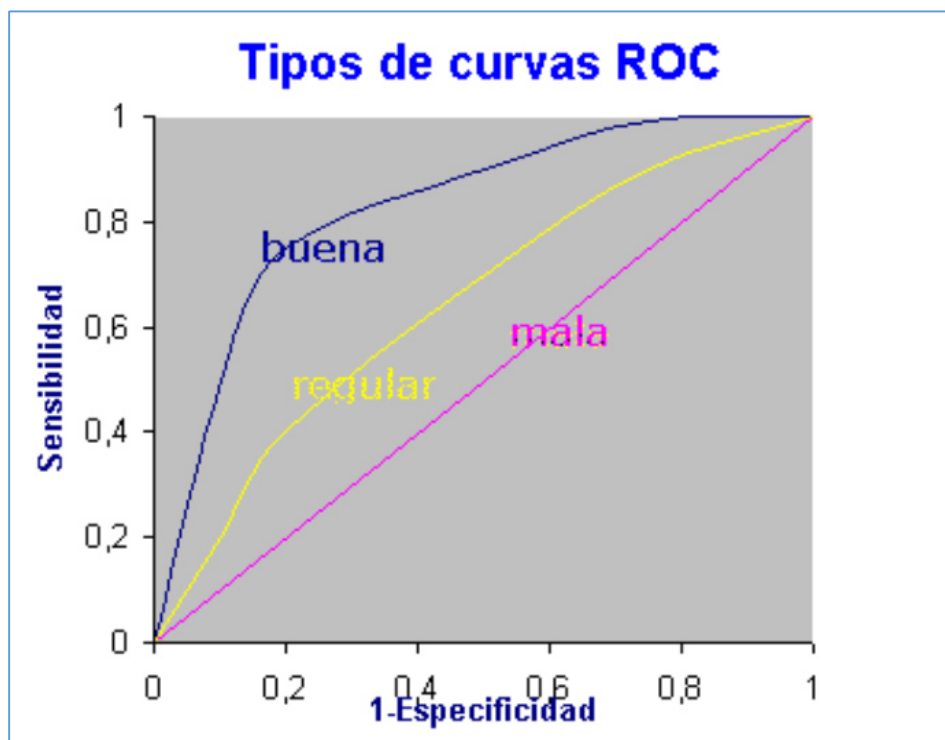


Gráfico 16: Curva ROC (Fuente: HRC-Comunidad Madrid)

A modo de guía para interpretar las curvas ROC (Gráfico 16) se han establecido los siguientes intervalos para los valores de AUC:

- [0.5]: Es como lanzar una moneda.
- [0.5, 0.6): Test malo.
- [0.6, 0.75): Test regular.
- [0.75, 0.9): Test bueno.
- [0.9, 0.97): Test muy bueno.
- [0.97, 1): Test excelente.

El AUC oscila en valor del 0 al 1. Un modelo cuyas predicciones son un 100% incorrecta tiene un AUC de 0.0; otro cuyas predicciones son un 100% correcta tiene un AUC de 1.0

Se calculó el AUC para las dos técnicas mencionadas (Ver Tabla 5)

Modelo	AUC (AREA BAJO LA CURVA ROC)
Árbol Jerárquico	0.66
Regresión Logística	0.72

Tabla 5: Comparación de Rendimiento de Algoritmos (Fuente: Elaboración Propia)

Se seleccionó el modelo de Regresión Logística por lo siguiente:

- ✓ El AUC de la regresión logística de 0.72 es mayor al del árbol jerárquico, por lo cual tiene mejor rendimiento y un nivel de predicción mayor.
- ✓ Los resultados del modelo son muy explicativos, lo que nos permite conocer la fuerza de asociación de las variables independientes con el efecto estudiado.
- ✓ En la literatura revisadas la regresión logística es la más usada para problemas donde la variable a predecir es dicotómica.
- ✓ Este algoritmo es muy fácil de implementar en cualquier entorno productivo, en nuestro problema será en sql.



#### ***D. Elección de Variables***

En el punto F se describirá este tópico debido a que es un proceso iterativo.

#### ***E. Entrenamiento***

En el punto F se describirá este tópico debido a que es un proceso iterativo.

#### ***F. Prueba del Modelo***

Para llegar al modelo adecuado realizamos 7 iteraciones donde consideramos lo siguiente:

- ✓ **Elección de Variable:** La variable debe tener una buena correlación con la variable Target y debe ser entendible, es decir tener significado para el negocio.
- ✓ **Entrenamiento:** Se debe realizar el modelado sobre la base de entrenamiento
- ✓ **Prueba:** Se debe replicar el modelo realizado sobre la base Test y medir el Rendimiento con el indicador AUC.

En la Tabla 6 se puede observar los resultados de las iteraciones más relevantes proporcionadas por **TIMI**:

Iteración	Variables	Importancia	Correlación	AUC
1	Sexo	16.4%	1.6%	0.827
	Antigüedad	9.7%	3.7%	
	Mes Campaña	8.1%	1.1%	
	Uso Línea	5.9%	3.7%	
	Nro_cta_ahorro	5.4%	1.1%	
	Región	4.3%	3.7%	
	Frescura en Campaña	3.9%	1.8%	
	Saldo_Otro_Banco	3.5%	3.7%	
2	Sexo	18.0%	0.7%	0.836
	Region	10.6%	1.7%	
	Antigüedad	10.1%	1.8%	
	Max_mto_operacion_anual	9.1%	1.8%	
	Flg_Operacion_T2	7.4%	1.8%	
	Saldo_PLD_6M	4.2%	1.6%	
	Banco_Max_Linea	3.8%	2.7%	
	Linea_entidad_financiera	3.3%	0.7%	
3	Sexo	21.0%	0.7%	0.841
	Region	12.1%	1.7%	
	Prom_Sem_Mto_Opera	7.7%	1.6%	
	Nro_Prom_Sem_tenencia_tc	7.0%	1.1%	
	Frescura	6.4%	0.7%	

	Flg_Operacion_T2	5.0%	1.6%	
	Edad	4.8%	1.7%	
	Uso_linea_trimestral	4.5%	2.7%	
7	Monto_Ahorro_1M	20.6%	1.3%	0.877
	Sexo	12.9%	0.7%	
	Uso_linea_trimestral	9.2%	1.1%	
	Edad	7.7%	2.7%	
	Frescura	6.8%	1.6%	
	Antigüedad	6.1%	1.6%	
	Flg_Operacion_T2	5.0%	0.7%	
	Prom_Sem_Mto_Opera	4.5%	0.3%	
	Monto_Salud	4.4%	1.1%	

Tabla 6: Resultados de Iteraciones del Modelado (Fuente: Elaboración Propia)

En la Iteración 7, conseguimos que el AUC del modelo llegue a 0.88 en su poder discriminativo, este indicador es aceptable debido al poco volumen de clientes Target que conforman nuestra muestra de entrenamiento.

A continuación, en la Tabla 7 y Gráfico 17, se detalla, las variables, los coeficientes de cada variable, y el tipo de relación con la variable Target identificadas para el modelo seleccionado de la iteración 7.

Variables	Coeficientes
Constante	-0.0147
Monto Ahorro	
Menor igual a 80	0
Menor igual a 752	0.0011
Menor igual a 825 K	0.0097
Mayor a 825 K	0.0098
Sexo	
F(Femenino)	0
M(Masculino)	0.009
Uso Línea TC	
Menor a 5%	0.0019
Menor a 10%	0.007
Menor a 30%	0.008
Menor a 50%	0.012
Mayor a 50%	0.012
Edad	
Menor a 20	0.001
Menor a 32	0.002
Menor a 42	0.004
Menor a 65	0.006

Mayor a 65	0.006
Antigüedad	
Menor a 1m	0.005
Menor a 36m	0.0045
Menor a 90m	0.0040
Menor a 1316m	0.003
Mayor a 1316m	0.002
Flag Operación	
0 (No)	0.003
1 (Si)	0.004
Nro promedio TC SF	
Igual a 0	0.0045
Menor igual a 1	0.002
Menor igual a 2	0.0004
Menor igual a 3	0.0003
Menor igual a 12	0.0001
Mayor a 12	0.00009
Monto Operaciones TC/TD	
Menor igual a 84	0
Menor igual a 337	0.0018
Menor igual a 700	0.0019

Menor igual a 1.2k	0.0050
Menor igual a 2.5k	0.0057
Menor igual a 350k	0.0063
Mayor a 350k	0.0064
Monto Giro Salud	
Menor igual a 43	0
Menor igual a 109	0.0009
Menor igual a 195	0.0011
Menor igual a 334	0.0013
Menor igual a 500	0.00168
Menor igual a 157k	0.0031
Mayor a 157k	0.0031

Tabla 7: Coeficientes del Modelo de Regresión Logística (Fuente: Elaboración Propia)

### Seguro Protección Múltiple Importancia de Variables

	Variable	Sentido
1	Monto de ahorro (últm mes)	↑
2	Sexo	■
3	% de Uso de TC (últm 3 meses)	↑
4	Edad	↑
5	Antigüedad del cliente	↓
6	Flag de operaciones (últm mes)	■
7	Nro. promedio de TC SF (últm 6 meses)	↓
8	Monto promedio de operaciones con TC y TD	↑
9	Máximo monto de giro salud (últm año)	↑

Gráfico 17: Variables del Modelo (Fuente: Elaboración Propia)

En el anexo A podemos ver como se relacionan las 9 variables importantes para el modelo para obtener una mayor efectividad en campañas, lo que aumenta las ventas.

## 5. Evaluación

Como parte de la evaluación del Modelo, se realizó lo siguiente:

- ✓ Se ejecutó el modelo sobre la base enviada en campaña de los meses de mayo 15 - octubre 15, donde se pudo observar como la efectividad de campaña se duplicaba en los leads más propensos (donde el modelo detecto mayor probabilidad de aceptación y permanencia)

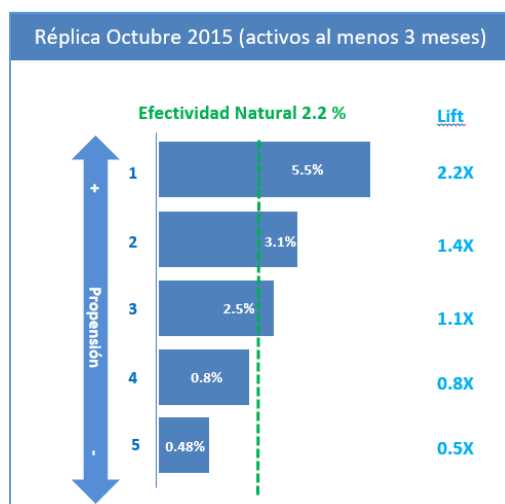


Gráfico 18: Comparación de la Efectividad Natural Vs el Modelo  
(Fuente: Elaboración Propia)

Con la base gestionada en Octubre15 (Gráfico 18), se observa que la efectividad de la base es de 2.2%, mientras que, si hubiésemos enviado solo leads de score de propensión altos como 1 y 2, tendríamos una efectividad de campaña por encima del 3%.

Así mismo se observa que el modelo replicado en los meses de campañas anteriores (Gráfico 19), mantiene los mismos resultados: duplica la efectividad de la campaña en el score más alto y cumple la escalera, es decir la efectividad de campaña siempre es más alta en los scores más altos y menor en los scores 4 y 5.

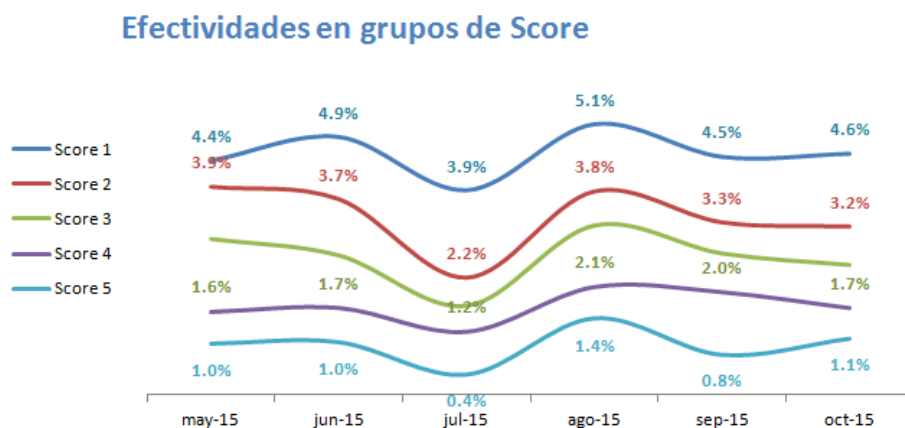


Gráfico 19: Efectividad según Score (Fuente: Elaboración Propia)

- ✓ Se desplegó campañas comerciales durante dos meses, lo cuales son denominados pilotos en donde se reta la utilización del modelo versus las lógicas de campañas sin modelo. Se obtuvo resultados satisfactorios los cuales son detallados en la sección 3.3.



## 6. Despliegue

El proceso de despliegue, utilización del modelo se detalla en la sección 3.2.5.

### 3.2.4. FUNDAMENTOS UTILIZADOS

**Metodología CrispDM:** Cross Industry Standard Process for Data Mining, es un modelo de proceso de minería de datos que describe una manera en la que los expertos en esta materia abordan el problema. Para implementar una tecnología en un negocio es necesaria una metodología. Estos métodos suelen venir de las experiencias propias y también de los procedimientos estándar más conocidos. En el caso de los proyectos de implementación de minería de datos una de las metodologías que ha tenido más apoyo de las empresas privadas y organismos públicos es CRISP-DM. (Moine, Haedo, & Gordillo, 2011)

CRISP-DM incluye un modelo y una guía, estructurados en seis fases, algunas de las cuales son bidireccionales, es decir que de una fase en concreto se puede volver a una fase anterior para poder revisarla, por lo que la sucesión de fases no tiene porqué ser ordenada desde la primera hasta la última. (Moine, Haedo, & Gordillo, 2011)

**Regresión Logística:** (Cuenca Ramallo, 2011) La Regresión Logística es un Algoritmo Supervisado y se utiliza para clasificación.

En estadística, la regresión logística es un modelo de regresión para variables dependientes, es útil para modelar la probabilidad de un evento ocurriendo como

función de otros factores. Es un modelo lineal generalizado que usa como función de enlace la función logit.

La regresión logística es usada extensamente en las ciencias médicas, sociales y otros, esto significa áreas del conocimiento humano donde las variables que se analizan son en su mayoría cualitativas.

El modelo de regresión logística se utiliza cuando estamos interesados en pronosticar la probabilidad de que ocurra o no un suceso determinado. Por ejemplo, a la vista de un conjunto de pruebas médicas, que una persona tenga una determinada enfermedad, o bien que un cliente devuelva un crédito bancario.

Se dice que un proceso es binomial cuando sólo tiene dos posibles resultados: “éxito” y “fracaso”, siendo la probabilidad de cada uno de ellos constante en una serie de repeticiones. (Ver Gráfico 20)

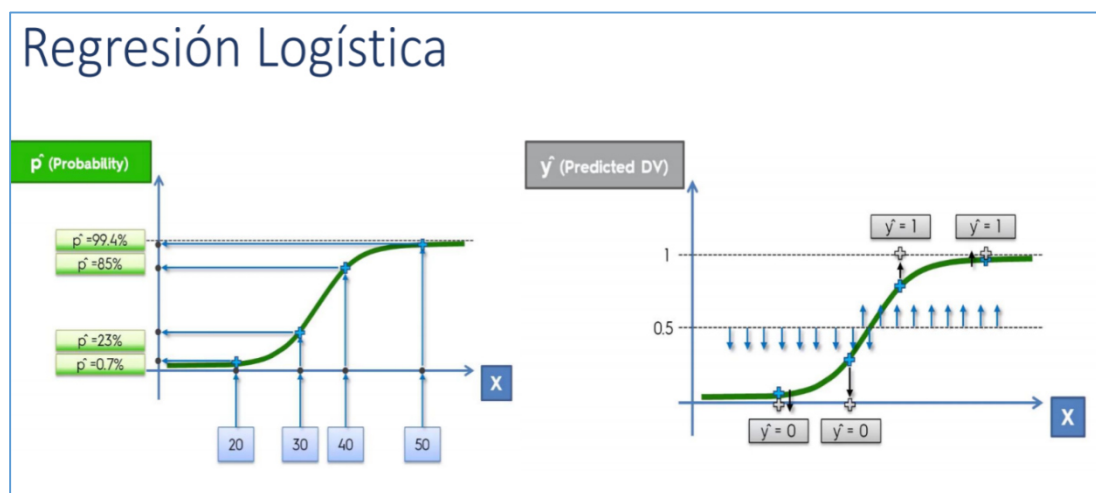


Gráfico 20: Algoritmo de Regresión Logística (Fuente: Programa Big Data y Analytics)

### 3.2.5. IMPLEMENTACION DE LAS AREAS, PROCESOS, SISTEMAS Y BUENAS PRÁCTICAS

El modelo de campañas se ha implementado en el área comercial, su utilización nos permite aumentar las ventas y reducir la fuga de clientes del producto, el detalle de los resultados se especificará en el ítem 3.3. A continuación se detalla el flujo de utilización del modelo en el despliegue de las campañas.

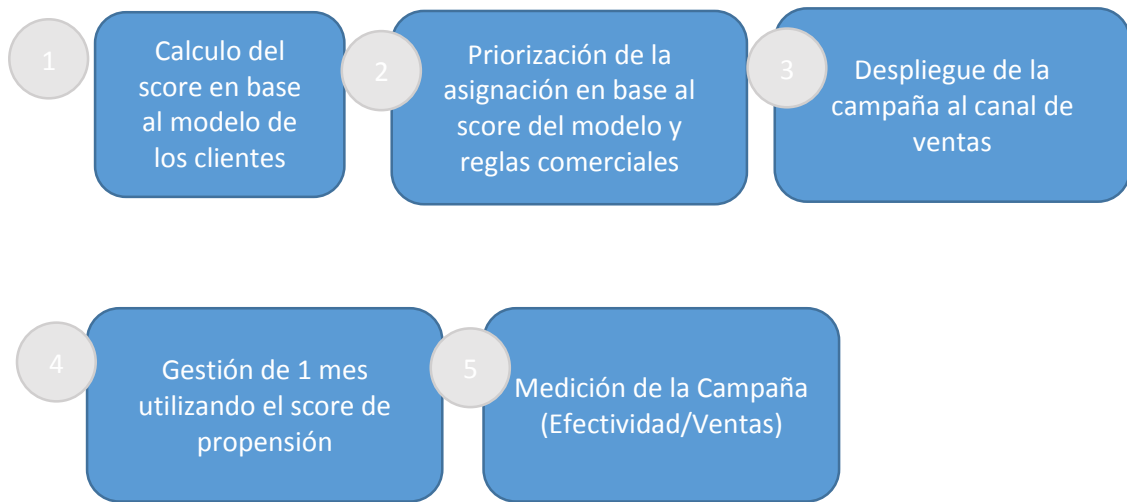


Gráfico 21: Flujo de Utilización del Modelo en las Campañas (Fuente: Elaboración Propia)

1. Cálculo del Score de los clientes (modelo): Mensualmente los 25 de cada mes se calcula la probabilidad de cada cliente para la captación del producto Seg. PM utilizando el modelo diseñado, y se entrega una marca de propensión por cada cliente del 1 al 5, donde 1 es mayor propenso a la compra y el 5 es el menos propenso.

2. Elaboración de la Campaña: En base a las reglas comerciales definidas en el Brief (Ver Anexo B) se elige el potencial que va ser enviado a los canales para ser gestionado.
3. Despliegue al canal: Se cuelga la campaña en el canal del proveedor, en este caso a un call center para que inicie la gestión.
4. Gestión: El canal tiene un plazo de 1 mes para gestionar la campaña, esta gestión se realiza en base al score de propensión, para que sus fuerzas se centren en los más propensos a adquirir el producto.
5. Medición: Cerrado el mes de gestión el equipo de conocimiento de clientes realiza una medición de las campañas donde se centra en indicadores como venta y efectividad, con estos resultados revisamos si los criterios tomados para la campaña deben corregirse o mantenerse.

### **3.3. EVALUACIÓN**

Con los indicadores del 2016 usando el modelo predictivo aumentamos nuestras ventas a 710 superando el KPI establecido, así como la efectividad de 2.8%. De igual manera el índice de cancelación se redujo a un 8% menor al que teníamos en el 2015 de 15%.

Con estos indicadores podemos afirmar que la solución nos permitió resolver los problemas detectados en el 2015, los cuales eran la disminución de ventas y el ratio de fuga de 15% después de 3 meses post venta. A continuación se detalla los indicadores de la evaluación:

#### **a. Prueba del Modelo Predictivo a través un Piloto Comercial**

Cuando se incorpora una variable nueva al flujo de campañas, como es la utilización de un modelo predictivo, el área realiza por 2 meses pilotos comerciales, los cuales consisten en enviar a gestionar 2 tipos de bases al canal. La primera base se elabora en base a las reglas comerciales sin la utilización del modelo y la segunda se escoge los prospectos a gestionar en base a la prioridad del modelo, para luego hacer un comparativo y evidenciar que la base priorizada tiene mejores resultados (ventas/efectividad).

En la realización del piloto participan 2 equipos:

**Equipo de Negocio:**

- Elabora el Brief Comercial, el cual es un documento oficial de la entidad financiera donde se especifica las reglas comerciales de la campaña, su objetivo, el canal donde se gestionará y el capacity. (Ver Anexo B)
- Crea el protocolo y speech de venta que debe usar al ejecutivo de venta al comunicarse con el cliente.
- Se encarga de comunicar al canal de venta, en este caso un call center cómo será la gestión de la base. Para el piloto, el canal debe priorizar en el barrido de llamadas a los clientes con mayor prioridad.

### **Equipo de Inteligencia Comercial:**

Para este piloto se siguió los siguientes pasos:

- Cálculo del Score de los clientes (modelo): Se calculó el score de los clientes de la entidad financiera (Ver Anexo C).
- Elaboración de la Campaña: En base a las reglas comerciales definidas en el Brief (Ver Anexo B) se elige el potencial que va ser enviado a los canales para ser gestionado, tanto la base priorizada con el modelo y la base a gestionar sin el modelo.
- Despliegue al canal: Se envía las dos bases al canal de call center, donde en la base priorizada por el canal colocamos el score para que prioricen su gestión
- Gestión: El canal tiene un mes para gestionar ambas bases con el mismo número de vueltas o barridos

- Medición: Finalizado el mes de Gestión, el equipo de conocimiento del cliente realiza la medición de las campañas pilotos. Los resultados se muestran a continuación.

Piloto (Marzo 16)			
Tipo Base	#Leads	Venta	Efectividad
Con Modelo	12,500	344	2.75%
Sin Modelo	12,500	270	2.16%

Piloto (Abril 16)			
Tipo Base	#Leads	Venta	Efectividad
Con Modelo	12,500	350	2.8%
Sin Modelo	12,500	280	2.2%

Gráfico 22: Indicadores de los Pilotos Comerciales Seg. PM (Fuente: Elaboración Propia)

En el Gráfico 22 se puede observar que las bases priorizadas con el modelo predictivo superan en 0.6% en efectividad a las bases regulares, de igual manera el número de pólizas desembolsadas. Al tener indicadores tan buenos el negocio aprueba la utilización del modelo para toda la campaña.

Ratio de Cancelaciones		
Tipo Base	Marzo 16	Abril 16
Con Modelo	8%	7.5%
Sin Modelo	15%	16%

Gráfico 23: Ratio de Cancelaciones de los Pilotos (Fuente: Elaboración Propia)

En el Gráfico 23. Se observa que la base priorizada con el modelo reduce en un 7.5% el ratio de fuga después de 3 meses post-venta. El ratio de fuga del modelo es de 8% en promedio menor al ratio de fuga promedio de 15% de la base sin priorización, lo cual reduce la problemática del ratio de fuga post venta de 3 meses.

#### **b. Comparativo de Ventas**

Se analiza las ventas de los últimos 4 meses de campaña del año 2015 y 2016



## Ventas del Seguro Protección Múltiple

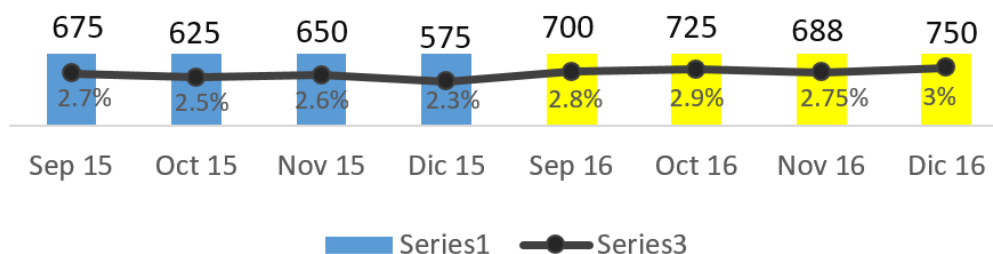


Gráfico 24: Ventas Seg. PM 2015 vs 2016 (Fuente: Elaboración Propia)

Para este comparativo se consideró las ventas en campañas de los últimos 4 meses del 2015 y 2016 para que sean comparables. Se puede observar que la efectividad promedio del 2015 es de 2.5% con ventas de 631 en promedio, mientras que en el 2016 con el uso del modelo la efectividad es de 2.85% con 710 ventas en promedio.

## **CAPÍTULO IV. REFLEXIÓN CRÍTICA DE LA EXPERIENCIA**

La implementación del modelo predictivo de Captación de Seguros de PM en la organización nos permitió resolver los problemas de las bajas ventas y el ratio de fuga de 15% del producto, llegando a cumplir con los objetivos de la organización. De esta experiencia comentaré los puntos más relevantes.

De la solución:

- El modelo predictivo con un poder discriminativo de 87.7%(auc), lo cual lo hace un modelo bueno, fue muy exitoso ya que nos permitió identificar el perfil correcto para las ventas de este producto, logrando incrementar las ventas y reducir la fuga.
- La técnica de regresión logística tuvo un poder discriminativo aceptable, pero podemos aumentar su rendimiento si utilizamos un poco más de historia (1 año como mínimo de data)
- La solución nos permitió identificar 9 variables principales para el modelo, lo cual no solo nos permitió entender la variable objetivo del estudio, sino que nos sirve de cimiento para aplicarlas en otras campañas de seguros para mejorar la segmentación e incrementar las ventas

En el Negocio:

- Después de los buenos resultados de los modelos que nos permitió conocer al público target que queríamos abordar, se desarrollaron 5 modelos para los productos de seguros tanto para la estrategia de retención como captación.
- Nos permitió direccionar mejor el esfuerzo de nuestros canales para reducir costos. Por ejemplo, a los clientes menos propensos (4,5) se dejó de gestionar por call center (Fuerza de Venta), y se utilizó en canal de email lo cual reduce los costos.
- Con el conocimiento del comportamiento del público objetivo a través de las variables más relevantes para el modelo, el negocio obtuvo las herramientas necesarias para hacer mejoras al producto y mejorar la oferta de valor por cada tipo de score.
- Generó mayor participación de los equipos de negocio o producto para la elaboración de los modelos, ya que parte importante del éxito de un modelo es que las variables deben tener sentido para el negocio permitiendo la explicación de la solución.

Como experiencia profesional:

- Participar en cada parte de la solución me permitió enriquecer mis conocimientos y poner en práctica los conocimientos teóricos sobre la elaboración de modelos predictivos.
- Como especialista de campañas dirigidas esta solución me dio las herramientas para mejorar las campañas comerciales de la entidad y el conocimiento sobre el perfil de los clientes del producto, lo cual me ayuda para mejorar la segmentación en las campañas.

- Pude aprender y poner en práctica conocimientos estadísticos, que son muy importantes para la interpretación de resultados y desarrollo del modelo.
- Aprendí el uso de una nueva herramienta Timi, al ser muy intuitiva y detallada con los resultados de cada modelado, permitió que especialistas con estudios diferentes a estadísticas e iniciantes en esta área puedan elaborar modelos de manera más ágil.

## **CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES**

### **5.1. CONCLUSIONES**

- La implementación del modelo predictivo para la captación de seguro PM y permanencia de 3 meses post venta, fue exitosa para la organización cumpliendo con sus dos objetivos: Reducción del ratio de fuga e incremento de las ventas. Comparando el 2015(campañas sin modelo) vs el 2016(campañas con modelo) se incrementó las de 631 pólizas en promedio a 710 pólizas, priorizando la gestión de los clientes con mayor probabilidad a la compra. En cuanto al ratio de fuga se redujo en 7% como se observa en los pilotos comerciales.
- Con el uso del modelo la organización llegó a cumplir sus indicadores del año 2016: Incrementar las ventas a 710 pólizas en promedio, aumentar la efectividad de campaña a 2.8% y reducir el ratio de fuga a 8%.
- Las 9 variables identificadas por el modelo nos permitieron conocer el perfil que necesitan este producto, lo cual nos permitió emplearlas en otros productos similares para la segmentación de campañas teniendo buenos resultados.
- El uso del indicador seleccionado para evaluar el modelo (Área bajo la curva Roc) fue muy acertado, el cual fue de 0.877, lo cual podemos verificarlos con los resultados alcanzados con la aplicación del modelo.

- Hoy en día el modelo sigue vigente siendo una de las variables principales usadas para la elaboración de las campañas comerciales para este producto. Los modelos predictivos están ayudando a las entidades financieras a orientar sus estrategias comerciales y lograr sus objetivos de ventas.

## 5.2. RECOMENDACIONES

- Probar con otros algoritmos de clasificación diferentes a regresión logística para evaluar si se consigue mayor poder de predicción.
- Se sugiere obtener una muestra mayor de 15% de la variable target para el entrenamiento del modelo, de esta manera se logra ser más finos en la comprensión del problema y nos permite poder discriminar con mayor exactitud los resultados positivos de los falsos positivos
- Se sugiere la participación activa del negocio para retroalimentarnos con su conocimiento sobre el comportamiento del producto, de esta manera logramos que las variables que escogemos para el modelo tengan significado y sentido para la resolución del problema. Hay muchos buenos modelos, pero pocos son los utilizados debido a que no son comprendidos por el negocio o los stakeholders.

## 5.3. GLOSARIO

- **BNET:** Banca por Internet de las Personas Naturales
- **ATM:** Cajeros automáticos
- **KPIS:** Indicador de medición

- **Datazo:** Plataforma para la carga de campañas en canales digitales
- **Segmentador de Colas:** Ticketera de la red de oficinas, el cual proporciona al cliente y al ejecutivo el orden de atención de los clientes.
- **PIC:** Plataforma Integral Comercial, donde se comunican las campañas comerciales a gestionar por la Red de Oficinas
- **Efectividad:** Es el indicador que nos permite evaluar cuan efectiva fue una campaña comercial, el cual se calcula de la división de las ventas realizadas entre los prospectos enviados a gestionar.
- **Campañas Comerciales:** Comprende un conjunto de acciones comerciales orientadas a un objetivo en específico como: captación de clientes, fidelización, retención y otros.

#### 5.4. FUENTES DE INFORMACIÓN

Las referencias de las fuentes de información son las siguientes:

Cuenca Ramallo, V. (2011). Regresión Logística. *Varianza*.

Hanley J.A., & McNeil B.J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*.

Moine, Haedo, & Gordillo. (2011). *Estudio Comparativo de metodologías para minería de datos*.

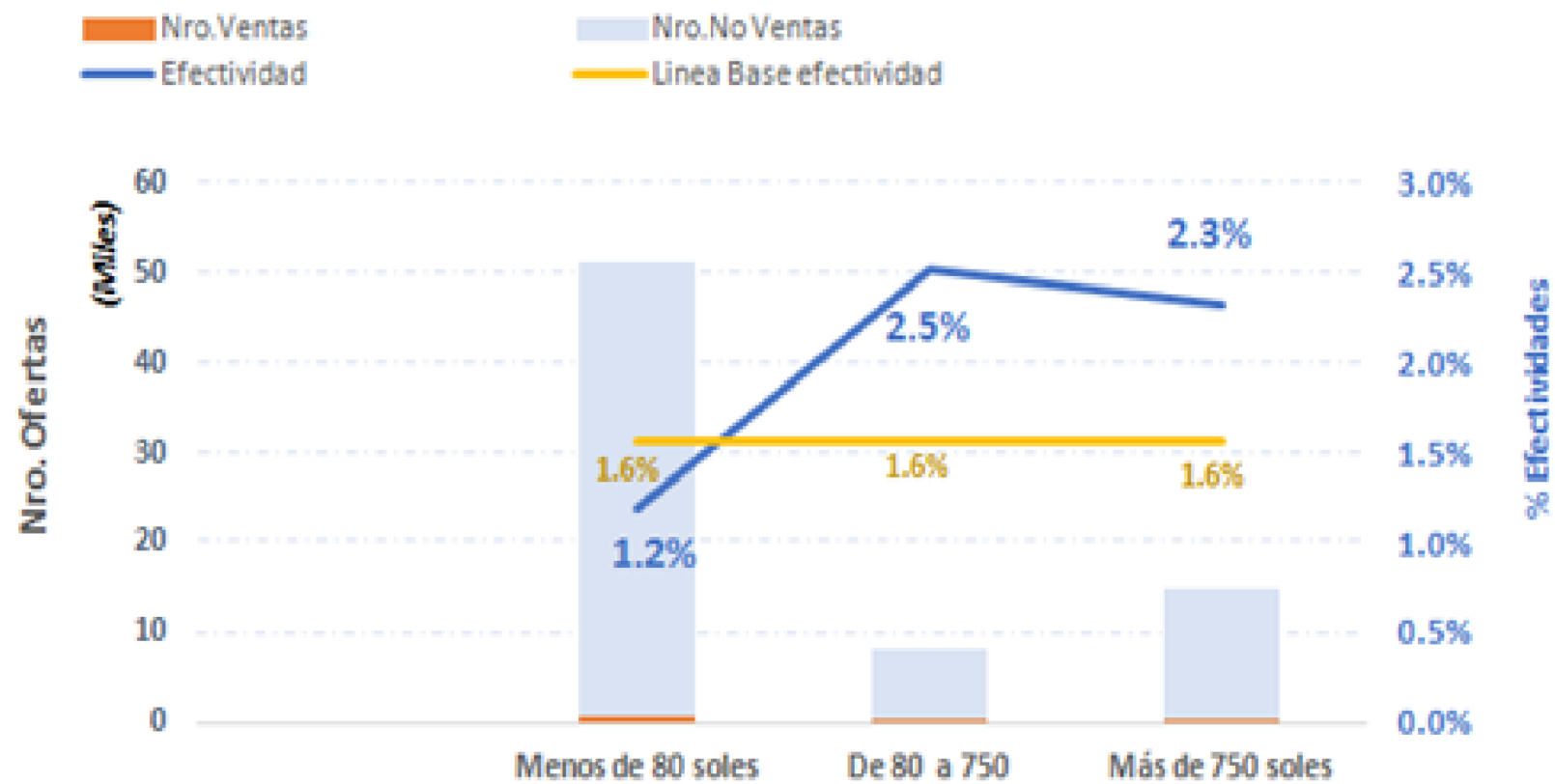
## **ANEXOS**

### **A. Variable del Modelos**

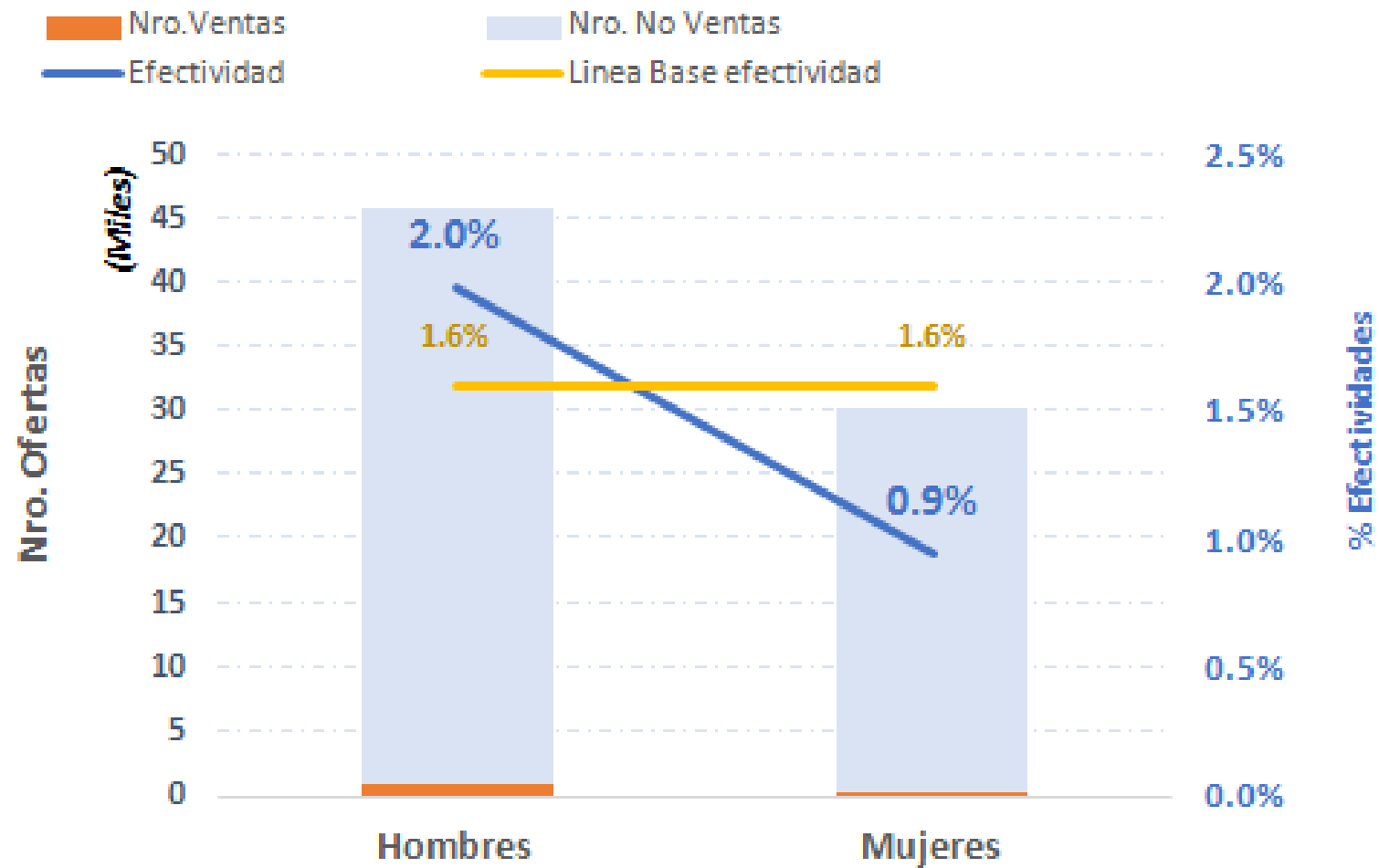
- **Los mejores prospectos de clientes de la entidad Financiera para ofrecer protección múltiple hombres con un monto de ahorro más de 80 soles**

**Monto de Ahorro**

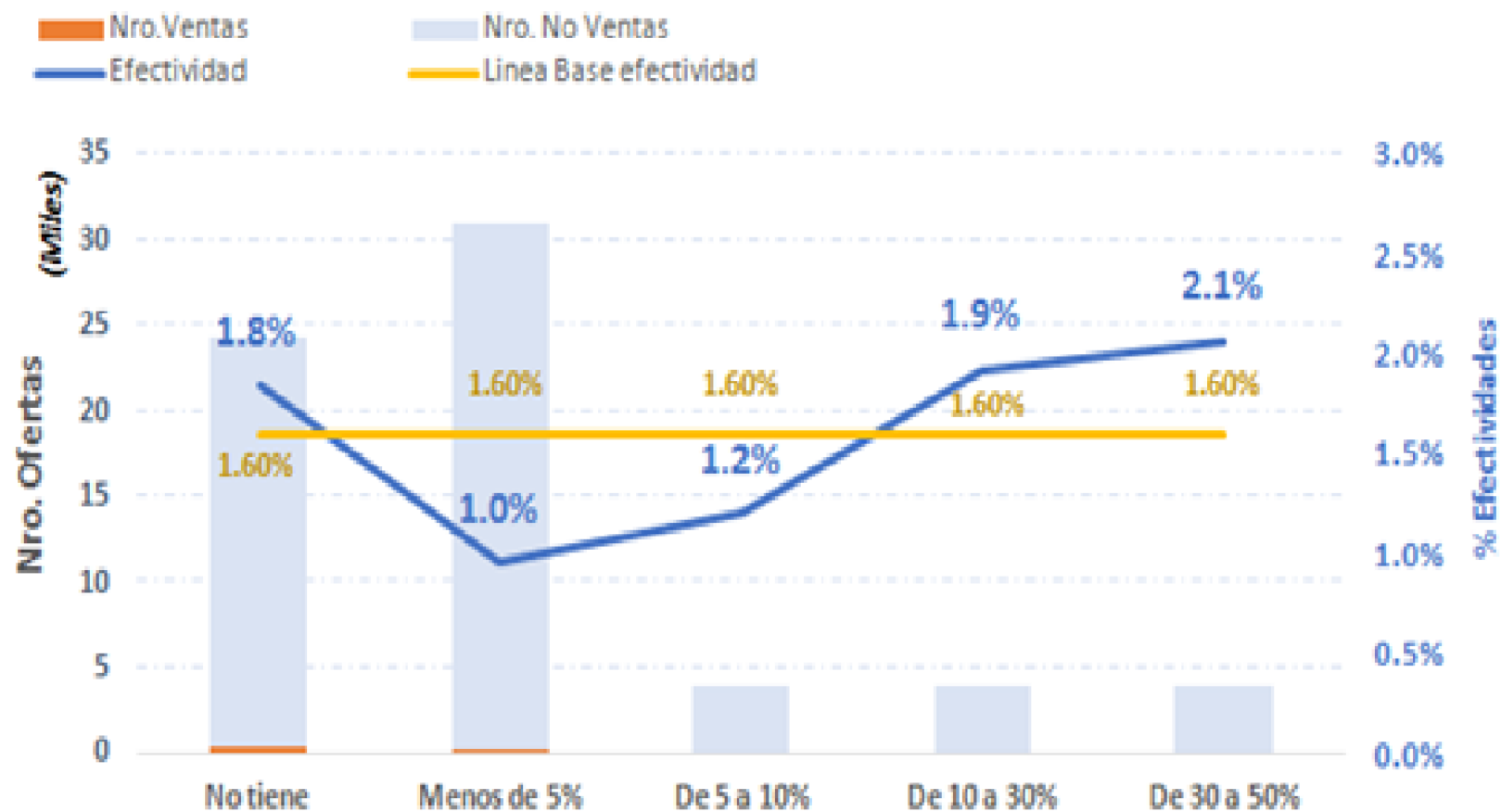




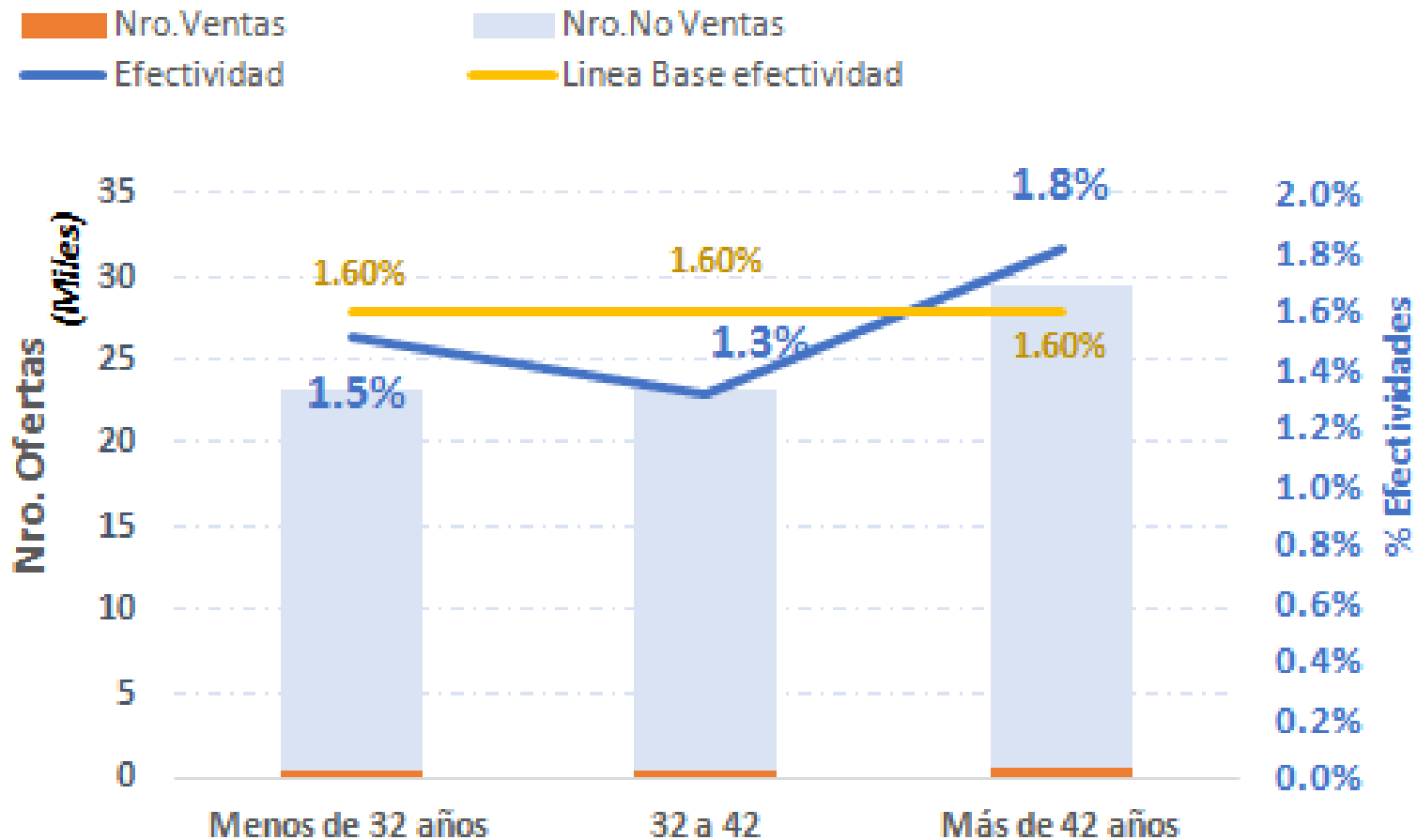
## Sexo



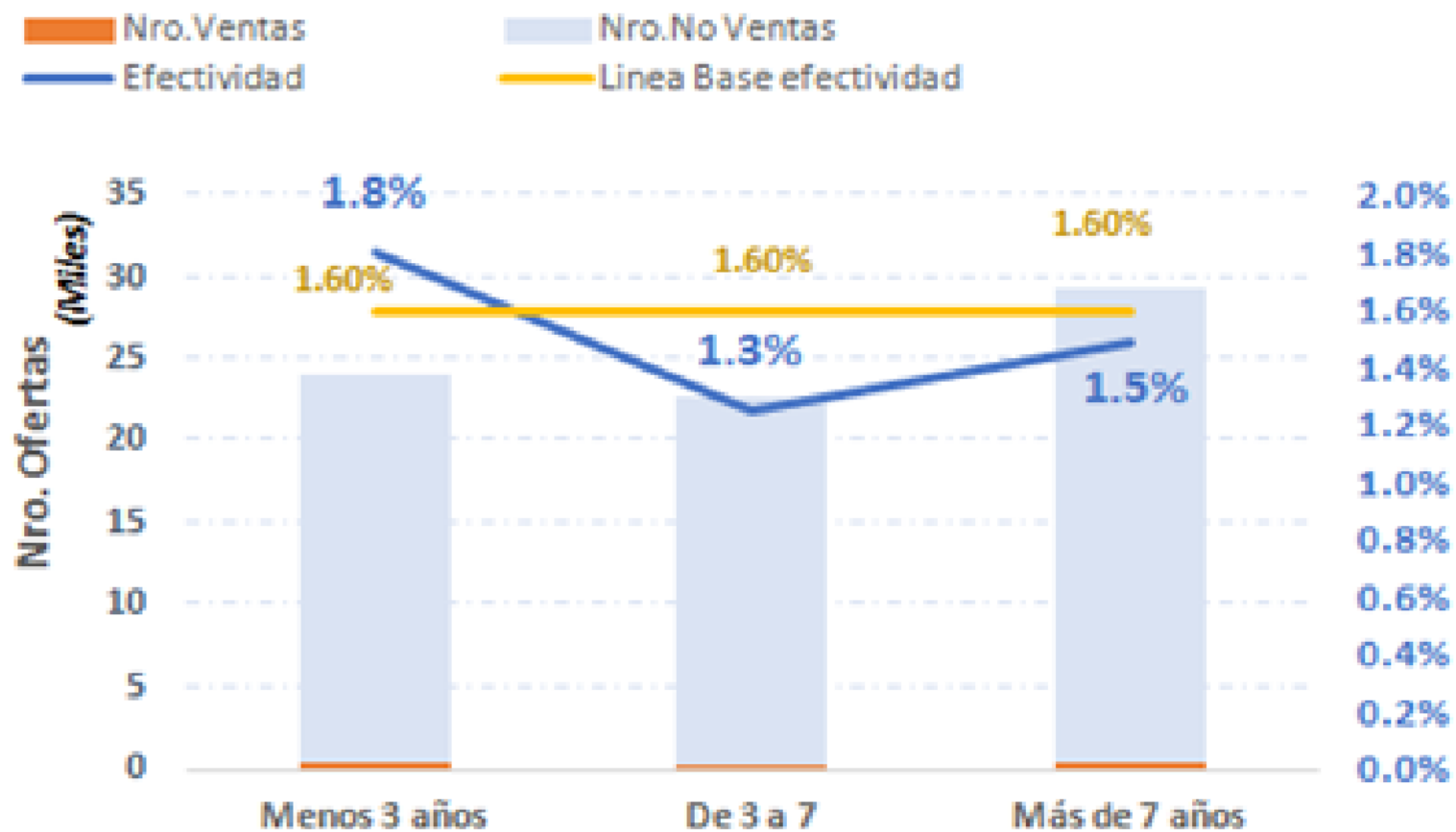
- Los clientes si tienen TC usan al menos 30% de la línea y edades mayores a 42 años



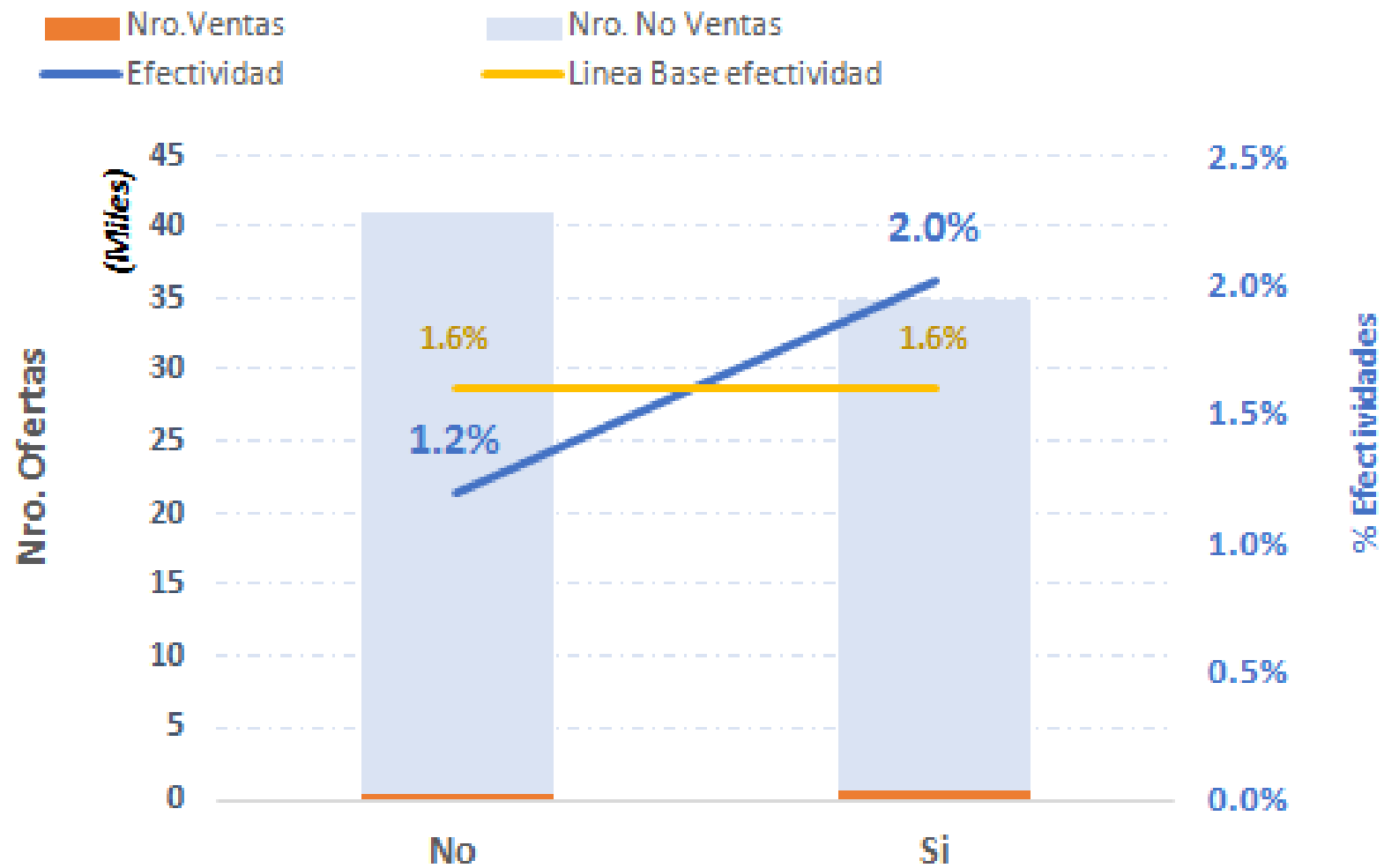
## Edad (años)



- Los clientes con menos 3 años en la entidad y que el último mes hayan realizado operaciones con su TC o TD



## Operaciones con TC o TD (últm mes)



**B. Brief Comercial**

Documento Oficial usado para definir las campañas comerciales que son desplegadas mes a mes.

## Antes del Modelo Predictivo

## Brief Comercial

## Campañas Dirigidas

Nombre de la Campaña	Seguro Protección Múltiple	Código IC	
Descripción	Seguro de invalidez total o permanente y de vida.		
Objetivo	Vender 700 seguros por producto	Responsable	Equipo de Seguros
Indicador de Medición	Efectividad Esperada 2.7%		

Tipo	Detalle
Filtros Generales	<ul style="list-style-type: none"> <li>- Black List.</li> <li>- Fallecidos.</li> <li>- Clasif. SBS Bueno</li> <li>- LPDP</li> <li>- Inelegibles</li> </ul> <p>Gracias no insista.</p>
Consideraciones	<p>'No incluir a:</p> <ul style="list-style-type: none"> <li>- Colaborador</li> <li>- Sin cta de ahorro ni TC</li> <li>- Clientes Segmento Altos</li> <li>- Sin Telefono de Contacto</li> <li>- Menor a 20 años y Mayor a 64 años</li> <li>- Clientes con saldo en cta ahorro promedio 3 meses menor a 600 y que en el ultimo mes el saldo sea &lt;600</li> <li>- Clientes que han sido enviados en los ultimos 3 meses en las Campañas (RH/PM, Seguro Veh)</li> <li>- Clientes con linea disponible TC &lt; 300</li> <li>- Clientes con tenencia de producto Renta Hospitalaria y Protección Multiple</li> <li>- Clientes con tarjeta VP</li> <li>- Clientes con TC activa menor a 2 meses</li> </ul> <p>- Clientes con RUC</p> <p>- Sólo teléfono fijo</p>



**Brief Comercial**  
Campañas Dirigidas

**Nombre de la Campaña**  
**Código IC**

Proteccion Múltiple

Canal	Filtros	Consideraciones
CALL CENTER		Capacidad: 25,000. Priorizar estas cantidades - Enviar 7% hasta 25 años - Enviar 40% provincia

### . Con el Modelo Predictivo

Brief Comercial			
Campañas Dirigidas			
<b>Nombre de la Campaña</b>	Seguro Protección Múltiple	<b>Código IC</b>	
<b>Descripción</b>	Seguro de invalidez total o permanente y de vida.		
<b>Objetivo</b>	Vender 700 seguros por producto	<b>Responsable</b>	Equipo de Seguros
<b>Indicador de Medición</b>	Efectividad Esperada 2.7%		

Tipo	Detalle
Filtros Generales	<ul style="list-style-type: none"> <li>- Black List.</li> <li>- Fallecidos.</li> <li>- Clasif. SBS Bueno</li> <li>- LPDP</li> <li>- Inelegibles</li> </ul> Gracias no insista.
Consideraciones	<p>'No incluir a:</p> <ul style="list-style-type: none"> <li>- Colaborador</li> <li>- Sin cta de ahorro ni TC</li> <li>- Clientes Segmento Altos</li> <li>- Sin Telefono de Contacto</li> <li>- Menor a 20 años y Mayor a 64 años</li> <li>- Clientes con saldo en cta ahorro promedio 3 meses menor a 600 y que en el ultimo mes el saldo sea &lt;600</li> <li>- Clientes que han sido enviados en los ultimos 3 meses en las Campañas (RH/PM, Seguro Veh)</li> <li>- Clientes con linea disponible TC &lt; 300</li> <li>- Clientes con tenencia de producto Renta Hospitalaria y Protección Multiple</li> <li>- Clientes con tarjeta VP</li> <li>- Clientes con TC activa menor a 2 meses</li> </ul> <p style="text-align: right;">- Clientes con RUC</p> <p style="text-align: right;">- Sólo teléfono fijo</p>

**Brief Comercial**  
Campañas Dirigidas

**Nombre de la Campaña**  
**Código IC**

Proteccion Múltiple

Canal	Filtros	Consideraciones
CALL CENTER		'Capacidad: 25,000. Priorizar estas cantidades Priorizar los score de propensión 1, 2. Si no llegas al capacity completar con el score 3.

Campañas

Pilotos

**Canales**

Comunicaciones

Datazo



**C. Score de los clientes utilizando el modelo predictivo.**

Results Messages				
	MES_CAMPANA	COD_PERSON	PROB_PROTMULTI	SCORE_PROTMULTI
1	201810	CLI273	0.0274988	2
2	201810	CLI275	0.0455294	1
3	201810	CLI277	0.0274988	2
4	201810	CLI278	0.0221512	2
5	201810	CLI273	0.0446161	1
6	201810	CLI274	0.0477497	1
7	201810	CLI276	0.0210768	2
8	201810	CLI270	0.0438353	1
9	201810	CLI271	0.0466281	1
10	201810	CLI274	0.0276234	2
11	201810	CLI277	0.0154445	3
12	201810	CLI275	0.0518692	1
13	201810	CLI273	0.0487291	1